Modeling and Predicting Popularity Dynamics via Reinforced Poisson Processes

Huawei Shen[1], Dashun Wang[2], Chaoming Song[3], Albert-László Barabási[4]
[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
[2]IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, USA
[3]Department of Physics, University of Miami, Coral Gables, Florida 33146, USA
[4]Center for Complex Network Research, Northeastern University, Boston, Massachusetts 02115, USA

Abstract

An ability to predict the popularity dynamics of *individual* items within a complex evolving system has important implications in an array of areas. Here we propose a generative probabilistic framework using a reinforced Poisson process to explicitly model the process through which individual items gain their popularity. This model distinguishes itself from existing models via its capability of modeling the arrival process of popularity and its remarkable power at predicting the popularity of individual items. It possesses the flexibility of applying Bayesian treatment to further improve the predictive power using a conjugate prior. Extensive experiments on a longitudinal citation dataset demonstrate that this model consistently outperforms existing popularity prediction methods.

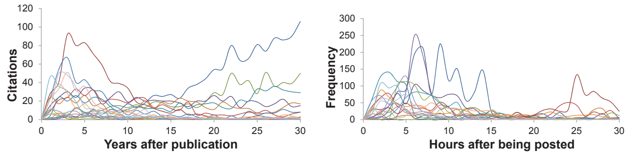
Introduction

Information explosion, from knowledge database to online media, places attention economy in the center of this era. In the heart of attention economy lies a competing process through which a few items become popular while most are forgotten over time (Wu and Humberman 2007). For example, videos on YouTube or stories on Digg gain their popularity by striving for views or votes (Szabo and Huberman 2010); papers increase their visibility by competing for citations from new papers (Wang, Song, and Barabási 2013); tweets or Hashtags in Twitter become more popular as being retweeted (Hong, Dan, and Davison 2011) and so do webpages as being attached by incoming hyperlinks (Ratkiewicz et al. 2010). An ability to predict the popularity of *individual* items within a dynamically evolving system not only probes our understanding of complex systems, but also has important implications in a wide range of domains, from marketing and traffic control to policy making and risk management. Despite recent advances of empirical methods, we lack a general modeling framework to predict the popularity of *individual* items within a complex evolving system.

Current models fall into two main paradigms, each with known strengths and limitations. One focuses on reproducing certain statistical quantities over an aggregation of items (Barabási 2005; Kempe, Kleinberg, and Tardos 2003; Backstrom et al. 2006; Dezso et al. 2006; Crane and Sornette 2008; Ratkiewicz et al. 2010). These models have been successful in understanding the underlying mechanisms of popularity dynamics. Yet, as they do not provide a way to extract item-specific parameters, these models lack predictive power for the popularity dynamics of individual items. The other line of enquiry, in contrast, treats the popularity dynamics as time series, making predictions by either exploiting temporal correlations (Szabo and Huberman 2010; Yang and Leskovec 2010; Lerman and Hogg 2010; Yan et al. 2011; Yu et al. 2012; Bao et al. 2013b) or fitting to these time series certain classes of functions (Bass 1969; Mahajan, Muller, and Bass 1990; Vu et al. 2011; Matsubara et al. 2012; Lerman and Hogg 2012; Gomez-Rodriguez, Leskovec, and Schölkopf 2013; Yang and Zha 2013). Despite their initial success in certain domains, these models are deterministic, modeling the popularity dynamics in a mean-field, if heuristic, fashion by focusing on the *average* amount of attention received within a fixed time window, ignoring the underlying arrival process of attentions. Indeed, to best of our knowledge, we lack a probabilistic framework currently to model and predict the popularity dynamics of individual items. The reason behind this is partly illustrated in Figure 1, suggesting that the dynamical processes governing individual items appear too noisy to be amenable to quantification.

In this paper, we model the stochastic popularity dynamics using reinforced Poisson processes, capturing simultaneously three key ingredients: fitness of an item, characterizing its inherent competitiveness against other items; a general temporal relaxation function, corresponding to the aging in the ability to attract new attention; and a reinforcement mechanism, documenting the well-known "rich-get-richer" phenomenon. The benefit of the proposed model is threefold: (1) It models the arrival process of individual attentions directly in contrast to relying on aggregated popularity time series; (2) As a generative probabilistic model, it can be easily incorporated into the Bayesian framework to account for external factors, hence leading to improved predictive power; (3) The flexibility in its choice of specific relaxation functions makes it a general framework that can be adapted to model the popularity dynamics in different domains.

Taking citation system as an exemplary case, we demonstrate the effectiveness of the proposed framework using a dataset peculiar in its longitudinality, spanning over 100 years and containing all the papers ever published by Amer-



(a) Citations (b) Hashtags

Figure 1: Stochastic Popularity dynamics. (a) 20 papers randomly selected from Physical Review during 1960s. (b) 20 Hashtags randomly selected from Twitter in 2012.

ican Physical Society. We find the proposed model consistently outperforms competing methods. Moreover, the proposed model is general. Hence it is not limited to predicting citations, but with appropriate adjustments will likely apply to other domains driven by competing processes.

Reinforced Poisson Process

The popularity dynamics of individual item d during time period $[0, T]$ is characterized by a set of time moments $\{t_i^d\}(1 \leq i \leq n_d)$ when each attention is received, where n_d represents the total number of attentions. Without loss of generality, we have $0 = t_0^d \leq t_1^d \leq \cdots \leq t_i^d \leq \cdots \leq t_{n_d}^d \leq T$. To model the arrival process of $\{t_i^d\}$, we consider two major phenomena confirmed independently in previous studies of population dynamics: (1) the reinforcement capturing the "rich-get-richer" mechanism, i.e., previous attention triggers more subsequent attentions (Crane and Sornette 2008); (2) the aging effect characterizing time-dependent attractiveness of individual items (Wang, Song, and Barabási 2013). Taken these two factors together, for an individual item d, we model its popularity dynamics as a reinforced Poisson process (RPP) (Pemantle 2007) characterized by the rate function $x_d(t)$ as

$$x_d(t) = \lambda_d f_d(t; \theta_d) i_d(t), \tag{1}$$

where λ_d is the intrinsic attractiveness, $f_d(t; \theta_d)$ is the relaxation function that characterizes the temporal inhomogeneity due to the aging effect modulated by parameters θ_d, and $i_d(t)$ is the total number of attentions received up to time t. From a Bayesian viewpoint, the total number of attentions $i_d(t)$ is the sum of the number of real attentions and the *effective* number of attentions which plays the role of prior belief. Here, we assume that all items are created equal and hence the effective number of attentions for all items has the same value, denoted by m. Therefore during the time interval between the $(i-1)$th and ith attentions, we have

$$i_d(t) = m + i - 1, \tag{2}$$

where $1 \leq i \leq n_d$. Accordingly, during the time interval between the n_dth attention and T, the total number of attention is $m + n_d$.

The length of time interval between two consecutive attentions follows an inhomogeneous Poisson process. Therefore, given that the $(i-1)$th attention arrives at t_{i-1}^d, the probability that the ith attention arrives at t_i^d follows

$$\begin{aligned} p_1(t_i^d | t_{i-1}^d) &= \lambda_d f_d(t_i^d; \theta_d)(m + i - 1) \\ &\quad \times e^{-\int_{t_{i-1}^d}^{t_i^d} \lambda_d f_d(t;\theta_d)(m+i-1)\mathrm{d}t}, \end{aligned} \tag{3}$$

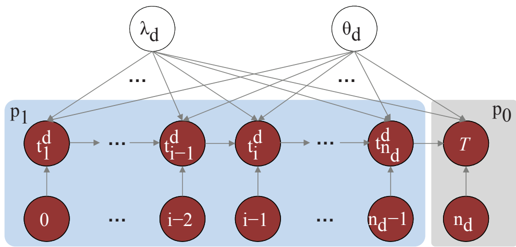


Figure 2: Graphical representation of the generative model for popularity dynamics via reinforced Poisson process.

and the probability that no attention arrives between $t_{n_d}^d$ and T is

$$p_0(T | t_{n_d}^d) = e^{-\int_{t_{n_d}^d}^{T} \lambda_d f_d(t;\theta_d)(m+n_d)\mathrm{d}t}. \tag{4}$$

Incorporating Eqs. (3) and (4) with the fact that attentions during different time intervals are statistically independent, the likelihood of observing the popularity dynamics $\{t_i^d\}$ during time interval $[0, T]$ follows

$$\begin{aligned} \mathcal{L}(\lambda_d, \theta_d) &= p_0(T | t_{n_d}^d) \prod_{i=1}^{n_d} p_1(t_i^d | t_{i-1}^d) \\ &= \lambda_d^{n_d} \prod_{i=1}^{n_d} (m + i - 1) f_d(t_i^d; \theta_d) \times \\ &\quad e^{-\lambda_d \left((m+n_d)F_d(T;\theta_d) - \sum_{i=1}^{n_d} F_d(t_i^d;\theta_d)\right)}, \end{aligned} \tag{5}$$

where $F_d(t; \theta_d) \equiv \int_0^t f_d(t; \theta_d)\mathrm{d}t$ and we have reorganized the terms on the exponent for simplicity. For clarity, we illustrate the proposed RPP model in the graphical representation (Figure 2).

By maximizing the likelihood function in Eq. (5), we obtain the most likely fitness parameter λ_d^* for item d in closed form:

$$\lambda_d^* = \frac{n_d}{(m + n_d)F_d(T; \theta_d^*) - \sum_{i=1}^{n_d} F_d(t_i^d; \theta_d^*)}. \tag{6}$$

The solution for θ_d^* depends on the specific form of relaxation function $f_d(t; \theta_d)$. We save the discussions about the estimation of θ_d^* for later.

Next we show that, with the obtained λ_d^* and θ_d^*, the model can be used to predict the expected number $c^d(t)$ of attention gathered by item d up to any given time t. Indeed, according to Eq. (1), for $t \geq T$, this prediction task is equivalent to the following differential equation

$$\frac{\mathrm{d}c^d(t)}{\mathrm{d}t} = \lambda_d f_d(t; \theta_d)(m + c^d(t)) \tag{7}$$

with the boundary condition $c^d(T) = n_d$. Solving this differential equation, we get the prediction function

$$c^d(t) = (m + n_d)e^{\lambda_d^*\left(F_d(t;\theta_d^*) - F_d(T;\theta_d^*)\right)} - m. \tag{8}$$

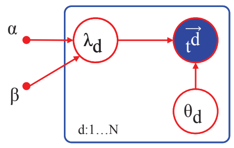


Figure 3: Probabilistic graphical model for reinforced Poisson process with conjugate prior.

Reinforced Poisson Process with prior

Maximum likelihood parameter estimation suffers from the overfitting problem for small sample size. For example, Eq. (6) gives $\lambda_d^* = 0$ when $n_d = 0$, and results in a null forecasting of future popularity, i.e., $c^d(t) = 0$ at any future time t. Moreover, the exponential dependency of $c^d(t)$ on λ_d^* in Eq. (8) leads to a large uncertainty in the prediction of $c^d(t)$. In this section, to overcome the drawback of the parameter estimation in Eq. (6), we adopt the Bayesian treatment for popularity prediction by introducing a conjugate prior for the fitness parameter λ_d, leading to a further improvement of the prediction accuracy of the proposed RPP model.

The likelihood function in Eq. (5) is a product of a power function and an exponential function of λ_d. Therefore, the conjugate prior for λ_d follows the gamma distribution

$$p(\lambda_d|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda_d^{\alpha-1}e^{-\beta\lambda_d}. \tag{9}$$

Note that this conjugate prior is the prior distribution of fitness parameters for all N items rather than for certain individual item. Hereafter, for convenience, we use $\vec{t^d} \equiv \{t_i^d\}$ to denote all the arrival time of attention gathered by item d. After introducing the conjugate prior, the graphical representation of model is depicted in Figure 3.

Using Bayes' theorem and combining Eqs. (5) and (9), we obtain the posterior distribution of λ_d

$$\begin{aligned} p(\lambda_d|\vec{t^d},\theta_d,\alpha,\beta) &= \frac{p(\vec{t^d}|\lambda_d,\theta_d)p(\lambda_d|\alpha,\beta)}{\int p(\vec{t^d}|\lambda_d,\theta_d)p(\lambda_d|\alpha,\beta)\mathrm{d}\lambda_d} \\ &= \frac{(\beta+X)^{\alpha+n_d}}{\Gamma(\alpha+n_d)}\lambda_d^{\alpha+n_d-1}e^{-(\beta+X)\lambda_d}, \end{aligned} \tag{10}$$

where $X \equiv (m+n_d)F_d(T;\theta_d) - \sum_{i=1}^{n_d} F_d(t_i^d;\theta_d)$.

With the obtained posterior distribution of λ_d, the expected number of attention $c^d(t)$, as shown in Eq. (8), can be predicted using its mean over the posterior distribution as

$$\begin{aligned} \langle c^d(t)\rangle &= \int c^d(t)p(\lambda_d|\vec{t^d},\theta_d,\alpha,\beta)\mathrm{d}\lambda_d \\ &= (m+n_d)\left(\frac{\beta+X}{\beta+X-Y}\right)^{\alpha+n_d} - m, \end{aligned} \tag{11}$$

where $Y \equiv F_d(t;\theta_d) - F_d(T;\theta_d)$. When $\beta \to \inf$, the prediction function reduces to a naive method, i.e., predicting that the popularity keeps constant in future. Eq. (11) is the Bayesian prediction function, predicting $c^d(t)$ using the posterior distribution of λ_d instead of using a single value of λ_d^* obtained by maximum likelihood estimation. Neither X, corresponding to empirical observations, nor Y, reflecting the rate difference in reinforced Poisson process, is in the exponent, indicating the robustness of this prediction function.

We now discuss how to determine the parameters α and β of prior distribution. Basically, the values of prior parameters could be tuned by checking the accuracy of prediction function with respect to prior parameters on so-called validation set. This means that we need to know the future popularity of some items to determine prior parameters. It is impractical in many scenarios where it is unrealistic to leverage future information for prediction.

One alternative solution is the fully Bayesian approach which introduces hyperprior for prior parameters. Although fully Bayesian approach is theoretically elegant, the inference of prior parameters is intractable in most cases. Approximation methods or Monte Carlo methods have to be adopted. As a result, the benefit of fully Bayesian approach is discounted by approximation gap in approximation methods or high computational cost of Monte Carlo methods.

In this paper, we determine the value of prior parameters by adopting maximum likelihood estimation with latent variable. Specifically, we choose the α and β values that maximize the following logarithmic likelihood function

$$\mathcal{L}(\alpha,\beta) = \sum_{d=1}^{N}\ln\int p(\vec{t^d}|\lambda_d)p(\lambda_d|\alpha,\beta)\mathrm{d}\lambda_d. \tag{12}$$

Here, θ_d is not explicitly written to keep the notation uncluttered. In sum, α and β are obtained according to

$$\frac{\partial\mathcal{L}(\alpha,\beta)}{\partial\beta} = \frac{N\alpha}{\beta} - \sum_{i=1}^{N}\lambda_d, \tag{13}$$

$$\begin{aligned} \frac{\partial\mathcal{L}(\alpha,\beta)}{\partial\alpha} &= N(\ln\beta - \phi_0(\alpha)) + \sum_{d=1}^{N}\ln\frac{\lambda_d}{\alpha+n_d} \\ &\quad + \sum_{d=1}^{N}\phi_0(\alpha+n_d), \end{aligned} \tag{14}$$

where ϕ_0 is the digamma function and the latent variable is

$$\lambda_d = \frac{\alpha+n_d}{\beta+(m+n_d)F_d(T;\theta_d) - \sum_{i=1}^{n_d}F_d(t_i^d;\theta_d)}. \tag{15}$$

Comparing Eq. (15) and Eq. (6), we can see that the fitness parameter λ_d is adjusted by prior parameters α and β.

Note that the parameters θ_d for all items are also determined by maximizing the likelihood function in Eq. (12). The calculation depends on the specific form of relaxation function $f_d(t;\theta_d)$, which is given in experiments on real dataset.

Experiments

In this section, we demonstrate the effectiveness of the proposed RPP model, with and without prior.

Experiment setup

Dataset. We conduct experiments on an excellent longitudinal dataset, containing all papers and citations published by American Physical Society between 1893 and 2009. We choose this dataset for two main reasons: (1) It covers an extended period of time, spanning 117 years, ideal for modeling and predicting temporal dynamics; (2) Treating papers as items, their popularity is well-defined, characterized by citations. Statistics about this dataset are shown in Table 1.

Relaxation function. When formalizing the model for popularity dynamics, we introduced a general relaxation function $f_d(t;\theta_d)$ and skipped the discussion of parameter θ_d. Here, when applying this model to a specific case, i.e., to citation system, we need to determine the specific form of the relaxation function as well as θ_d. Previous studies (Radicchi, Fortunato, and Castellano 2008; Wang, Song, and Barabási 2013) on citation dynamics suggest that the aging of papers is captured by a log-normal relaxation function

$$f_d(t;\mu_d,\sigma_d) = \frac{1}{\sqrt{2\pi}\sigma_d t}\exp\left(-\frac{(\ln t-\mu_d)^2}{2\sigma_d^2}\right), \quad (16)$$

a common relaxation function, which is also observed in other domains such as messages in microblogging networks (Bao et al. 2013a).

For item d with log-normal relaxation function, θ_d is replaced by parameters μ_d and σ_d, which can be calculated by maximizing the logarithmic likelihood $\mathcal{L}$ in Eq. (12) and Eq. (5) for the proposed RPP model with and without prior, respectively. In this paper, we maximize logarithmic likelihood using optimization methods which leverage gradients

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\mu_d} &= \frac{1}{\sigma_d}\left\{\sum_{i=1}^{n_d}\left[\tau_i^d-\lambda_d\phi(\tau_i^d)\right]\right.\\ &\left.+\lambda_d(n_d+m)\phi(\tau^d)\right\}, \quad (17)\\ \frac{\partial\mathcal{L}}{\partial\sigma_d} &= \frac{1}{\sigma_d}\left\{\sum_{i=1}^{n_d}\left[\tau_i^d*\tau_i^d-\lambda_d\tau_i^d\phi(\tau_i^d)\right]\right.\\ &\left.+\lambda_d(n_d+m)\tau^d\phi(\tau^d)-n_d\right\}, \quad (18)\end{aligned}$$

where ϕ is the probability density function of standard normal distribution, $\tau_i^d \equiv (\ln t_i^d-\mu_d)/\sigma_d$ and $\tau^d \equiv (\ln T-\mu_d)/\sigma_d$. Therefore, we can use Eqs. (17) and (18) together with Eqs. (13) and (14) to maximize the logarithmic likelihood in Eq. (12) for the RPP model with prior, together with Eq. (6) to maximize the likelihood in Eq. (5) for the RPP model without prior.

Baseline models and evaluation metrics. We compare the RPP model with three widely-used models for popularity prediction: the classic autoregression (AR) method (Box, Jenkins, and Reinsel 2008), the linear regression method of logarithmic popularity (SH) (Szabo and Huberman 2010), and the WSB model (Wang, Song, and Barabási 2013), which is equivalent to the proposed RPP model without prior when the log-normal relaxation function is adopted. We adopt two standard measurements as evaluation metrics:

Table 1: Basic statistics of dataset.

Journal	#Papers	#Citations	Period
PRSI	1, 469	668	1893-1912
PR	47, 941	590, 665	1913-1969
PRA	53, 655	418, 196	1970-2009
PRB	137, 999	1, 191, 515	1970-2009
PRC	29, 935	202, 312	1970-2009
PRD	56, 616	526, 930	1970-2009
PRE	35, 944	154, 133	1993-2009
PRL	95, 516	1, 507, 974	1958-2009
RMP	2, 926	115, 697	1929-2009
PRSTAB	1, 257	2, 457	1998-2009
PRSTPER	90	0	2005-2009
Total	463, 348	4, 710, 547	1893-2009

- Mean Absolute Percentage Error (*MAPE*) measures the average deviation between predicted and empirical popularity over an aggregation of items. Denoting with $c^d(t)$ the predicted number of citations for a paper d up to time t and with $r^d(t)$ its real number of citations, we obtain the MAPE over N papers

$$MAPE = \frac{1}{N}\sum_{d=1}^{N}\left|\frac{c^d(t)-r^d(t)}{r^d(t)}\right|.$$

- *Accuracy* measures the fraction of papers correctly predicted for a given error tolerance ϵ. Hence the accuracy of popularity prediction on N papers is

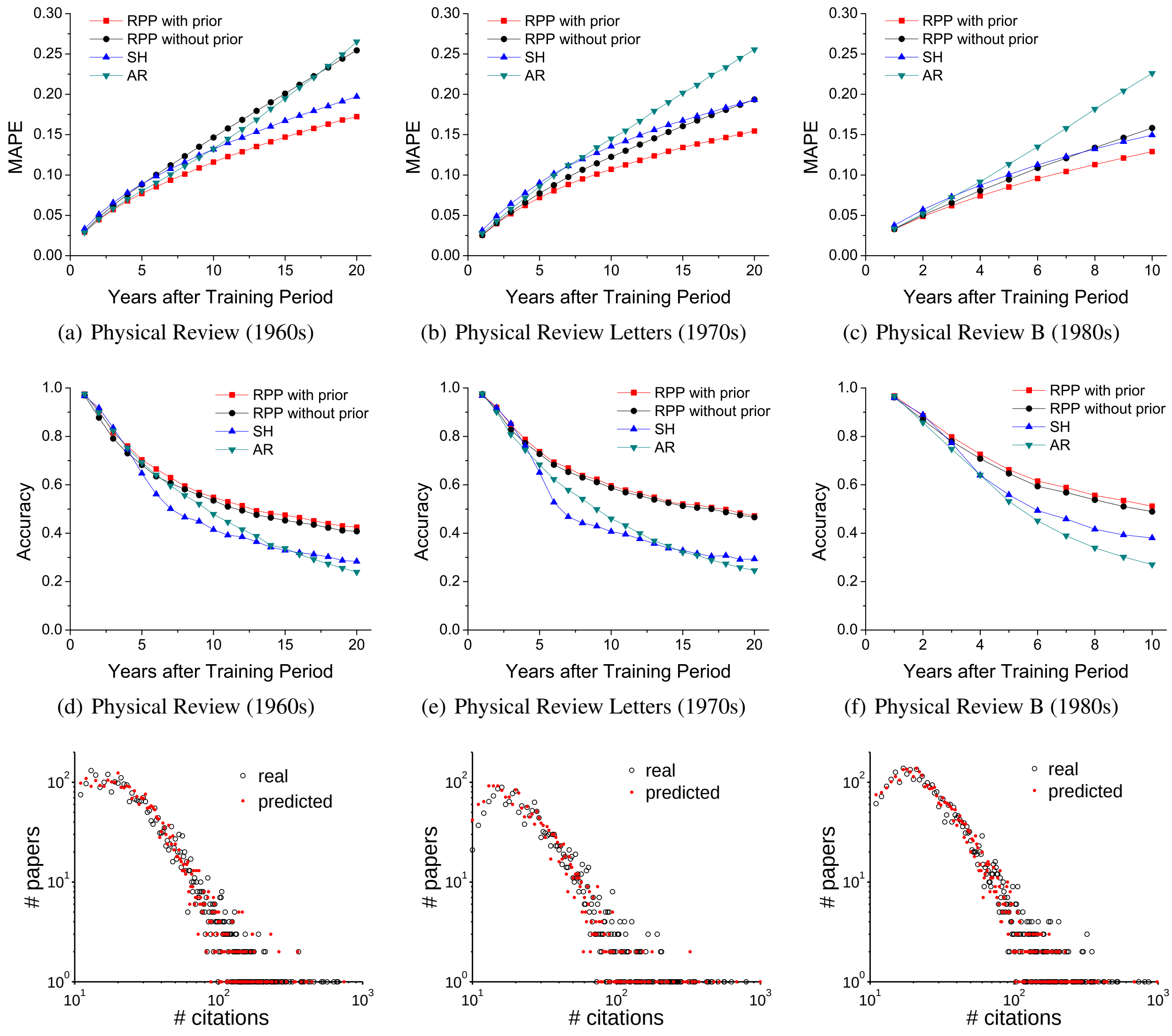
$$\frac{1}{N}\sum_{d=1}^{N}|\{d:\left|\frac{c^d(t)-r^d(t)}{r^d(t)}\right|\le\epsilon\}|.$$

We set the threshold $\epsilon = 0.1$ in this paper.

Experiment Results

In this section, we report two sets of experiments: (1) We compare the predictive power of RPP model with other competing methods, finding that RPP consistently outperforms other models; (2) We perform detailed analysis to understand the factors that could affect the performance of RPP model, including the length of training period, the effective number of attention, and the prior parameters.

Popularity prediction. We evaluate the prediction results on three collections of papers: (a) papers published in Physical Review (PR) from 1960 to 1969; (b) papers published in Physical Review Letters (PRL) from 1970 to 1979; (c) papers published in Physical Review B (PRB) from 1980 to 1989. These samples vary in timeframes and scopes, spanning three decades and covering three types of journals. Using papers with more than 10 citations during the first five years after publication, we compare the RPP model with and without prior against the AR and SH models. The number of papers in the three collections is 3242, 2017 and 3732, respectively. The training period is 10 years and we predict the citation counts for each paper from the 1st to 20th year after the training period. For collection (c), we predict the citation counts up to the 10th year after training period due to the cutoff year (2009). We set the parameter $m = 30$ for now, corresponding to the typical number of references for



(a) Physical Review (1960s) (b) Physical Review Letters (1970s) (c) Physical Review B (1980s)

(d) Physical Review (1960s) (e) Physical Review Letters (1970s) (f) Physical Review B (1980s)

(g) Physical Review (1960s) (h) Physical Review Letters (1970s) (i) Physical Review B (1980s)

Figure 4: The performance comparison in popularity prediction.

a paper, leaving the effect of varying m on the performance of RPP model for later discussions.

We find the RPP model, proposed in this paper, achieves higher accuracy than the AR and SH models (Figure 4). Yet in absence of prior it only exhibits modest performance in terms of MAPE, indicating that the RPP model without prior performs well on most papers but has rather large errors on a handful of papers. This is caused by its exponential dependence on the fitness parameter that sometimes yields overfitting problem when maximum likelihood parameter estimation is adopted. This problem is nicely avoided by incorporating conjugate prior for the fitness parameter, documented by the fact that the RPP model with prior consistently outperforms the other three methods on all collections.

The superiority of the RPP model with prior, compared to the AR and SH models, increases with the number of years after the training period. This improvement is rooted in the methodological advantage: the RPP model is a generative probabilistic model that explicitly models the arrival process of attentions, while the two baseline models only capture the correlation between early popularity and future popularity, linearly or logarithmically. In addition, the reinforced Poisson process could model the “rich-get-richer” phenomenon in popularity dynamics and thus could characterize the logarithmic correlation between early popularity and future popularity. Therefore, when compared with the AR method, the superiority is more obvious than being compared with the SH method. This is because the AR method works linearly while the SH method works in a logarithmic manner.

The RPP models with and without prior are trained only on the popularity dynamics during training period while the training of the AR and SH models depend on the knowledge of future popularity dynamics. When training these two models, we employ the leave-one-out technique which uses

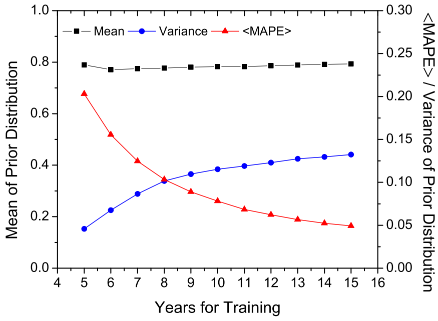


Figure 5: Effect of training period length.

Table 2: Effect of the number of conceived attention.

m	Mean (α/β)	Variance (α/β^2)	$\langle$MAPE$\rangle$
10	1.467	0.193	0.0762
20	1.005	0.150	0.0776
30	0.783	0.115	0.0781
40	0.647	0.091	0.0784
50	0.554	0.074	0.0785

all papers except the target paper for prediction. Yet, in most cases, it is unrealistic to know future popularity dynamics when training the model, limiting their applications in real scenarios.

Finally, being a generative model, the RPP model is able to reproduce the citation distribution. Indeed, as shown in Figure 4 (g-i), the distribution of citations predicted by the RPP model with prior matches very well with that of real citations on all studied collections, indicating that the RPP model can also be used to model the global properties of citation system.

Analysis of relevant factors. The superior predictive power in the RPP model with prior raises an interesting question: what are the possible factors that affect its predictive power? In this section, we study a number of factors which could affect the performance of the RPP model with prior. Hereafter, we use $\langle$MAPE$\rangle$ to denote the average MAPEs for predictions from the 1st to 10th year after training period. The training period is 10 years except when we discuss the effect of varying training period length. The parameter m is set to be 30 except when we discuss the effect of changing m.

First, we study the prediction accuracy of the RPP model with prior by varying training period. Experiments are conducted on the collection of papers published in Physical Review from 1960 to 1969. As shown in Figure 5, $\langle$MAPE$\rangle$ decreases as the training period increases. Hence increasing the training period improves the prediction accuracy. However, the rate at which $\langle$MAPE$\rangle$ diminishes slows down quickly, indicating the marginal gain of increasing training period. We also find that the mean of prior distribution stays almost constant as the length of training period increases from 5 years to 15 years, indicating the expected fitness parameter learned by the RPP model is robust against varying training period. At the same time, a longer training period could reduce the role of prior in prediction, partly explaining the role of prior in overcoming the overfitting problem, as demonstrated by the increasing variance in the prior distributions with the length of training period.

Table 3: Prediction accuracy over four decades.

Period	α	β	α/β	$\langle$MAPE$\rangle$
1950s	4.237	4.061	1.043	0.075
1960s	4.759	4.440	1.072	0.084
1970s	6.130	4.924	1.245	0.111
1980s	10.706	5.379	1.990	0.120

Second, we investigate the effect of parameter m, i.e., the effective number of attention by conducting experiments on the paper collection (a). Intuitively, m balances the strength in the reinforcement mechanism. Indeed, as shown in Table 2, the mean and variance of the prior distribution decay with m, demonstrating these parameters are mainly determined by papers with fewer citations. We also find that decreasing m reduces $\langle$MAPE$\rangle$, indicating that the disparity in citations is captured appropriately by the reinforcement mechanism in our model, as a larger m implies a weaker role of the reinforcement mechanism. Taken together, Table 2 confirms that the reinforcement mechanism is crucial to modeling popularity dynamics in citation system.

Finally, we use papers published in Reviews of Modern Physics (RMP) to illustrate the change of prior parameter α and β over four decades and their influence on the prediction accuracy of the RPP model with prior. As shown in Table 3, the mean of prior distribution (i.e., α/β) increases with the increasing magnitude of both α and β over the four decades. This indicates that the expected citations for papers in this prestigious journal steadily increases in the second half of the 20th century. Meanwhile, the $\langle$MAPE$\rangle$ of the RPP model also increases. Hence it becomes more difficult to predict the citations of these papers, as a result of the increasingly skewed distribution of citations.

Conclusions

Taken together, we presented a general framework to model and predict popularity dynamics based on a reinforced Poisson process. This model incorporates three key ingredients of popularity dynamics: the fitness parameter characterizing intrinsic attractiveness, the temporal relaxation function explaining the aging effect in attracting new attentions, and the reinforcement mechanism corresponding to the "rich-get-richer" effect in popularity dynamics. Being a generative probabilistic framework, it explicitly models the stochastic process of gaining popularity for each item, in contrast to existing deterministic approaches. We developed optimization methods to train the proposed RPP model with and without priors. The RPP model with prior allows us to apply the Bayesian treatment, resulting in more robust and accurate predictions for popularity dynamics. We empirically validate our model on an excellent longitudinal dataset on citations, spanning over one hundred years, demonstrating its clear advantages over competing methods.

Acknowledgments

This work was funded by the National Basic Research Program of China (973 Program) under grant numbers 2014CB340401, 2013CB329602, the National High-tech

R&D Program of China (863 Program) under grant number 2014AA015103, and the National Natural Science Foundation of China (Nos. 61202215, 61232010). This work was also partly funded by the Beijing Natural Science Foundation under grant number 4122077. DW, CS, ALB are supported by Lockheed Martin Corporation (SRA 11.18.11), the Network Science Collaborative Technology Alliance is sponsored by the U.S. Army Research Laboratory under agreement W911NF-09-2-0053, Defence Advanced Research Projects Agency under agreement 11645021, and the Future and Emerging Technologies Project 317 532 "Multiplex" financed by the European Commission.

References

Backstrom, L.; Huttenlocher, D.; Kleinberg, J.; and Lan, X. 2006. Group formation in large social networks: Membership, growth, and evolution. In *KDD'06*, 44–54.

Bao, P.; Shen, H. W.; Chen, W.; and Cheng, X. Q. 2013a. Cumulative effect in information diffusion: empirical study on a microblogging network. *PLoS ONE* 8(10):e76027.

Bao, P.; Shen, H. W.; Huang, J.; and Cheng, X. Q. 2013b. Popularity prediction in microblogging network: A case study on sina weibo. In *WWW'13*, 177–178.

Barabási, A. L. 2005. The origin of bursts and heavy tails in human dynamics. *Nature* 435(7039):207–211.

Bass, F. M. 1969. A new product growth for model consumer durables. *Management Science* 15(5):215–227.

Box, G. E. P.; Jenkins, G. M.; and Reinsel, G. C. 2008. *Time Series Analysis: Forecasting and Control*. Wiley, 4th edition.

Crane, R., and Sornette, D. 2008. Robust dynamic classes revealed by measuring the response function of a social system. *PNAS* 105(41):15649–15653.

Dezso, Z.; Almaas, E.; Lukács, A.; Rácz, B.; Szakadát, I.; and Barabási, A. L. 2006. Dynamics of information access on the web. *Physical Review E* 73:066132.

Gomez-Rodriguez, M.; Leskovec, J.; and Schölkopf, B. 2013. Modeling information propagation with survival theory. In *ICML'13*.

Hong, L.; Dan, O.; and Davison, B. D. 2011. Predicting popular messages in twitter. In *WWW'11*, 57–58.

Kempe, D.; Kleinberg, J.; and Tardos, E. 2003. Maximizing the spread of influence through a social network. In *KDD'03*, 137–146.

Lerman, K., and Hogg, T. 2010. Using a model of social dynamics to predict popularity of news. In *WWW'10*, 621–630.

Lerman, K., and Hogg, T. 2012. Using stochastic models to describe and predict social dynamics of web users. *ACM Transactions on Intelligent Systems and Technology* 3(4):62.

Mahajan, V.; Muller, E.; and Bass, F. M. 1990. New product diffusion models in marketing: A review and directions for research. *The Journal of Marketing* 54:1–26.

Matsubara, Y.; Sakurai, Y.; Prakash, B. A.; Li, L.; and Faloutsos, C. 2012. Rise and fall patterns of information diffusion: Model and implications. In *KDD'12*, 6–14.

Pemantle, R. 2007. A survey of random processes with reinforcement. *Probability Surveys* 4:1–79.

Radicchi, F.; Fortunato, S.; and Castellano, C. 2008. Universality of citation distribution: toward an objective measure of scientific impact. *PNAS* 105(45):17268–17272.

Ratkiewicz, J.; Fortunato, S.; Flammini, A.; Menczer, F.; and Vespignani, A. 2010. Characterizing and modeling the dynamics of online popularity. *Physical Review Letters* 105(15):158701.

Szabo, G., and Huberman, B. A. 2010. Predicting the popularity of online content. *Communications of the ACM* 53(8):80–88.

Vu, D. Q.; Asuncion, A. U.; Hunter, D. R.; and Smyth, P. 2011. Dynamic egocentric models for citation networks. In *ICML'11*, 857–864.

Wang, D.; Song, C.; and Barabási, A. L. 2013. Quantifying long-term scientific impact. *Science* 342:127–132.

Wu, F., and Humberman, B. 2007. Novelty and collective attention. *PNAS* 104(45):17599–17601.

Yan, R.; Tang, J.; Liu, X.; Shan, D.; and Li, X. 2011. Citation count prediction: learning to estimate future citations for literature. In *CIKM'11*, 1247–1252.

Yang, J., and Leskovec, J. 2010. Modeling information diffusion in implict networks. In *ICDM'10*, 599–608.

Yang, S. H., and Zha, H. 2013. Mixture of mutually exciting processes for viral diffusion. In *ICML'13*, 1–9.

Yu, X.; Gu, Q.; Zhou, M.; and Han, J. 2012. Citation prediction in heterogeneous bibliographic networks. In *SDM'12*, 1119–1130.