Modeling and Quantifying the Forces Driving Online Video Popularity Evolution

Jiqiang Wu, Yipeng Zhou, Dah Ming Chiu
Email: jqwu@tcl.com; yipeng.job@gmail.com; dmchiu@ie.cuhk.edu.hk

Abstract—Video popularity is an essential reference for optimizing resource allocation and video recommendation in online video services. However, there is still no convincing model that can accurately depict a video’s popularity evolution. In this paper, we propose a dynamic popularity model by modeling the video information diffusion process driven by various forms of recommendation. Through fitting the model with real traces collected from a practical system, we can quantify the strengths of the recommendation forces. Such quantification can lead to characterizing video popularity patterns, user behaviors and recommendation strategies, which is illustrated by a case study of TV episodes.

I. INTRODUCTION

Video popularity is an essential reference for optimizing resource allocation and video recommendation in online video services. Over the past few years researchers have studied several aspects of video popularity such as popularity distributions [1], evolution patterns [2], prediction [3][4], models [5][6] and so on. However, there is still no convincing model that can accurately depict a video’s popularity evolution. Such a dynamic model is useful not only for predicting popularity evolution, but also for characterizing user behaviors and content providers’ recommendation strategies.

Large-scale online video providers provide millions of videos such that users heavily rely on recommendations to watch their favored videos (besides ad-hoc searching). As a result, a video’s popularity evolution is largely affected by how it is recommended to users. Direct and word-of-mouth (WOM) recommendations are the two major mechanisms which are repeatedly observed and discussed by previous works [7][8][6]. Direct recommendation means that videos are exposed to users typically through websites’ front pages, TV promotion channels, advertisements, etc and WOM means that videos are shared through various social networks such as Facebook, Twitter, BBS forums, emails and even offline social networks like colleges and relatives.

The two recommendation mechanisms result in an information diffusion process. By modeling this process we can derive the incremental user population of watching a video and thus derive the video’s popularity evolution. However, the complication of such a model lies in the fact that practical recommendation resources are limited. For example, because of limited positions in front-page promotion, existing videos can only be recommended for a certain period as new videos are constantly brought in. So far there is little knowledge of how the two recommendations drive popularity dynamics with limited resources, which will be discussed by this study.

On the other hand, given a video’s observed popularity evolution, we can fit it to the model to reveal and quantify its spreading process and the driving forces of recommendation. Such quantification will provide a systematic approach to characterizing a dynamic video system by a set of insightful parameters. Both academia and industry can benefit from this approach as it can be used to detect latent user behaviors, identify inadequately recommended videos and evaluate recommendation strategies. We will illustrate it through a case study of TV episodes in Section [7].

II. DEFINITIONS AND ASSUMPTIONS

Video popularity is measured by view count throughout this paper. Without loss of generality, we discuss the popularity evolution for a particular video. Let \( x(t) \) be the cumulative popularity evolution function, i.e the number of users who have watched the video up to time \( t \). Then the (instant) popularity at time \( t \) is \( x'(t) \). In our model, all recommendations are classified as either direct recommendation or WOM recommendation. The recommendation made by content creators, providers and any other entity for their own benefits, is regarded as direct recommendation; while the recommendation made by users based on their own interests is regarded as WOM recommendation. We define the intrinsic attractiveness \( q \) of a video as the probability that the video is favored by any user. As users learn the video’s existence through recommendations, only an average fraction \( q \) of them will watch the video.

To simplify the analysis, we make the following assumptions. Firstly, users never replay a video. This assumption is supported by some previous work [2][8]. Secondly, users make independent decision to select videos. It may not be true for related videos. We will introduce a technique in Sec. [7] to tackle this issue. Thirdly, for a given video, the total potential user population that can be recommended is fixed. In practice, this assumption holds for a relatively short period, e.g. several months. Finally, users watch the video immediately after they learn its existence and are interested in it. Some previous work assume there is a user reaction process. However, for popular videos this process can be ignored, as an approximation [6].

III. MODEL

The information diffusion process may be driven by both direct recommendation and WOM recommendation, but it is complicated to involve both of them in a single process. A previous work [6] has shown that for a given video either direct
or WOM recommendation plays the main driving role. Therefore we study two simplified information diffusion processes: DModel driven by direct recommendation, and WModel driven by WOM recommendation.

Without loss of generality, we model a particular video’s popularity evolution by assuming users make independent video selection as stated in Sec. III. We create a fluid epidemic model to depict its information diffusion process. Let $\Delta t$ denote the elapsed time since the video is born (made available online). At time $t$, $S(t)$ is the set of users who do not know the existence of the video. $X(t)$ is the set of users who know the existence of and are interested in the video, and $Y(t)$ is the set of users who know the existence of but are not interested in the video. Let $s(t)$, $x(t)$ and $y(t)$ be the cardinality of the sets $S(t)$, $X(t)$ and $Y(t)$, respectively. Under the assumption of fixed population $N$, we have $s(t) + x(t) + y(t) = N$, and approximately $\frac{x(t)}{y(t)} = \frac{x}{N}$. Finally, let $v(t) = x'(t)$ be the dynamic view count, i.e. the (instant) popularity at time $t$.

### A. DModel

Because of limited recommendation resource in practice, it is impossible that the video information can be diffused to all users instantly. Therefore we define $\alpha$ as the direct recommendation rate, i.e. the rate that users flow from set $S(t)$ to $X(t) \cup Y(t)$. Note that the total potential user population $N$ is typically a very large number, e.g. several tens of millions. Thus we can create a fluid model as follows:

\[
\begin{align*}
x'(t) &= \alpha q s(t), \\
y'(t) &= \alpha (1-q) s(t),
\end{align*}
\]

with initial condition $x(0) = y(0) = 0$ and $s(0) = N$.

Together with the equation $s(t) + x(t) + y(t) = N$, we can derive that

\[
\begin{align*}
x(t) &= q N (1 - e^{-\alpha t}) \\
x'(t) &= \alpha q N e^{-\alpha t}
\end{align*}
\]

There is no restriction on direct recommendation period in the above process. In practice, as new videos are constantly brought into the system, each existing video is only promoted for a certain period, depending on the estimated video popularity, user feedback, arrival rate of new videos and so on. Thus we introduce another parameter $t_e$ as the time when the direct recommendation aborts. After time $t_e$, no new user will join in $X(t)$ and then $x'(t)$ will decrease to 0 very fast with rate $\gamma \gg \alpha$.

Summarizing the above discussion, we derive the video popularity evolution driven by direct recommendation as

\[
v(t) = \begin{cases} \\
\alpha q N e^{-\alpha t}, & 0 < t \leq t_e \\
\alpha q N e^{-\alpha t_e} \cdot e^{-\gamma (t-t_e)}, & t > t_e
\end{cases}
\]

### B. WModel

In the WModel we split the timeline into time slots. Let $D(t)$ denote the set of new users who join the set $X(t)$ in time slot $t$, and let $\Delta x(t) = x(t) - x(t-1)$ be the cardinality of $D(t)$. In time slot $t$, the users in $D(t)$ perform two operations:

- a) finish watching the video and thus generate $\Delta x(t)$ view count, and
- b) recommend the video to other users merely in this time slot. In other words, users recommend a video immediately after they finish watching it and the recommendation period only lasts one time slot. This is consistent with the reality. User attentions are often attracted by new videos such that no users will keep recommending a video for a long period. The granularity of time slot can be one day or one week depending on the scenario.

Let $\beta$ be the average number of friends a new user will recommend the video, normalized by $N$. Then the WOM recommendation process is a discrete time model:

\[
\begin{align*}
\Delta x(t+1) &= \beta q \Delta x(t)s(t), \\
y(t+1) &= \frac{x(t+1)(1-q)}{q},
\end{align*}
\]

The interpretation of the above equations is as follows. The new users joining the set $X(t+1)$ in time slot $t+1$, i.e. the $\Delta x(t+1)$ users in $D(t+1)$, are attracted by WOM recommendation made by the $\Delta x(t)$ users in $D(t)$. $\beta N$ is the average number of users who will be recommended by a user and $\frac{s(t)}{N}$ is the probability that the video information is unknown for any user. Thus, $\beta N \times \frac{s(t)}{N} \times q$ (only a fraction $q$ are interested) is the number of new users contributed by a user in $D(t)$. Given $\Delta x(t)$ users recommending the video, we have $\Delta x(t+1) = \beta q \Delta x(t)s(t)$.

It is hard to get the closed-form solution of $x(t)$ or $v(t)$. To obtain its theoretical insights, we consider an approximate continuous model by letting $x''(t) \approx \frac{\Delta x(t+1) - \Delta x(t)}{\Delta t}$ and $x'(t) \approx \frac{\Delta x(t)}{\Delta t}$. Then Eq. \ref{eq:final} becomes

\[
x''(t) = x'(t)[\beta q s(t) - 1]
\]

In order for the WOM to proceed, we should have an initial condition $x(0) = x_0 > 0$, $x_0$ is the number of seeds and we allow $x_0 < 1$. The period with $x(t) < 1$ indicates how long it takes for the first user to discover the video.

Given the fixed total user population $s(t) + \frac{x(t)}{q} = N$ (recall $\frac{x(t)}{q(t)} = \frac{g(t)}{1-q}$) and initial conditions that $x(0) = x_0$, $x'(0) \approx x'(1) = x(0) = \beta q x_0 (N - x_0 q^{-1})$, we can get

\[
x(t) = \frac{x_1 + g(t)x_2}{1 + g(t)}
\]

Here $x_1 = q N + \frac{q-1}{q}$, $x_2 = q N + \frac{q-1}{q}$, $g(t) = \frac{x_1 - x_0}{x_0 - x_2} e^{-\varphi t}$ and $\varphi = \sqrt{(\beta q N - 1)^2 - (\beta x_0 - 1)^2 + 1}$. The detailed derivation is in [9].

The solution of $x(t)$ is complicated for deriving the evolution of $v(t)$. Instead, we focus on analyzing the limit of $x(t)$. Clearly, as $t$ approaches infinity, $g(t)$ approaches 0 and $x(t)$ approaches $x_1$. In other words, the limit of $x(t)$ depends on the value of $x_1$, which should never exceed $q N$. The gap $q N - x_1$ is essential as it indicates whether the video has been sufficiently recommended.

\footnote{This means a user recommends the video to random $\beta N$ users among the population without keeping track of who have watched the video.}
Proposition 1: If $\beta \geq \frac{2}{qN+x_0}$, the video information can be diffused to all users; otherwise, only a subset of users know the video information.

The video information is diffused to all users if and only if $x_1 \geq qN$. It is not difficult to derive the condition $\beta < \frac{2}{qN+x_0}$ by letting $x_1 < qN$. The detailed proof is omitted. From Proposition 1 we can see that either small $\beta$ or $x_0$ can result in insufficient recommendation.

In the following discussion, we focus on the case where $x_1 < qN$. We are interested in how $x_0$ and $\beta$ affect $x_1$, i.e. the number of users who finally watch the video.

**Proposition 2:** If $\beta < \frac{2}{qN+x_0}$, the final user population $x_1$ increases as a concave function of $x_0$.

This proposition can be proved from $\frac{\partial^2 x_1}{\partial x_0^2} = \frac{-\beta(\beta qN-1)}{2\beta^2 x_0^3} < 0$ and $\frac{\partial x_1}{\partial x_0} = \frac{-\beta(\beta qN-1)^2 + 1}{\beta^2 x_0^2} < 0$. Fig. 1 shows how $x_0$ affects final population $x_1$ under different values of $\beta$. We can see that $x_1$ only increases from 5000 to 15000 as $x_0$ increases from 1000 to 5000 when $\beta x_0 = 1.67$, indicating the inefficiency for diffusing information by increasing $x_0$.

**Proposition 3:** There exists a $\frac{1}{qN+x_0} < \theta < \frac{2}{qN+x_0}$, such that if $\beta \leq \theta$ the final user population $x_1$ increases as a convex function of $\beta$; while if $\theta \leq \beta < \frac{2}{qN+x_0}$ the final user population $x_1$ increases as a concave function of $\beta$.

The proof is in [9]. Fig. 2 shows how $x_1$ increases as $\beta$ increases under different values of $x_0$. There exists a threshold such that if $\beta$ is larger than the threshold, it is very effective to lift the final user population $x_1$. However, in real world, it is much more difficult for video providers to control $\beta$, because $\beta$ is mainly determined by individual users.

C. Discussion

Although we have derived the DModel and WModel based on different driving forces of recommendation, for a given video it is unknown which model fits it best. If the view count trace of the video can be obtained, the type of information diffusion process is decided by the model with smaller fitting error. In addition, through fitting the video traces with the theoretical models, we can quantify the strength of each recommendation force, which is helpful for evaluating recommendation strategies and observing the user behavior. The details will be presented in Sec. V.

The assumption that a video’s information is diffused either by direct recommendation or WOM recommendation may not be valid, because both forces can simultaneously drive the process. However, it is reported in [6] that for most videos only one force plays the major role. Therefore, we can merely consider one single force to simplify the information diffusion process. With the simplification, the limitation of recommendation resources (e.g. $t_e$ and short WOM period) can be studied, which is a significant advantage of our model over the model proposed in [6] that analyzed multiple forces simultaneously.

IV. Evaluation

In this section, we verify the DModel and WModel by fitting them with the view count traces collected from a practical system. Each video will be fitted with both models, but only the one with smaller error is used for evaluation. For convenience, we refer to the model with better fitting as BModel.

The normalized mean square error(NMSE) is used as the metric to evaluate fitting error, which is defined as

$$NMSE = \frac{1}{T} \sum_{t=1}^{T} \frac{(\hat{v}_t - v_t)^2}{\bar{v}_t^2},$$

where $\hat{v}_t$ and $v_t$ represent the $t^{th}$ day’s view count calculated by the model and collected from the real system, respectively. Note that NMSE has been normalized by the square of the real trace’s average view count so that we can compare cases with different total view count.

A. Dataset

Instead of evaluating all videos, we focus on the videos from the four most important types: Movie, TV, News and Music Video (MV). We collected all videos of the four types that are uploaded between September 1, 2014 to January 31, 2015, from the system’s viewing records. Each viewing record contains the following information: time, user id and video id. For each video, we collected its daily view count trace for six months. The videos with less than 1000 total view count are removed from evaluation because their traces are likely to be driven by occasional views hence are too noisy. Removing cold videos does not affect the generality of this study since their view count takes less than 5% of the total view count. The resulting dataset contains 1469 movies, 9705 TV episodes, 30720 News and 4736 MVs.

Before fitting the models with the trace dataset, we need to figure out the total user population for each video type. We cannot simply take the number of total user id as the population because it would include occasional viewers. Instead, we consider the relatively active users who have watched a fair number of videos. Thus, for each video type we rank all users according to their view count in increasing order. Then the users with few view count are excluded until 25% total view count are removed, along with their contributed views to the
video. The number of remaining users varies from around 44 millions to around 81 millions. Although this is a heuristic rule to exclude occasional users, one does not have to exactly get the value $N$, which is mainly used to normalize the other parameters $\alpha$, $\beta$, etc. As long as the same $N$ is used for a video type, the comparison is fair for videos within the same type.

### B. Model fitting

![Fig. 2: Fitting error distributions of the four video types.](image)

The Levenberg-Marquardt algorithm, a common algorithm for finding least square error, is used to search for the optimal parameters when fitting the models with trace data. The NMSE for each video is calculated by Eq. 9 with the optimal parameters. We use the EvoModel proposed in [6] as the benchmark and compare the cumulative distribution functions (CDF) of NMSE. The results are plotted in Fig. 2. Note that the NMSE of BModel is the smaller NMSE of DModel and WModel. Through model fitting, we automatically classify the videos into two groups: d-recommended videos and w-recommended videos.

From Fig. 2 we can see that for Movie and TV, BModel achieves better performance than EvoModel (the two CDF curves of Movie almost overlap), while EvoModel slightly outperforms BModel for News and MV. We explain these results as follows. EvoModel analyzes multiple forces simultaneously without restricting each recommendation resource ($t_r$ and short WOM period). According to [6], the information of most movies and TV episodes is diffused either by direct recommendation or WOM recommendation. Thus BModel can achieve better fitting results by incorporating the limitation of recommendation resources. However, for News and MV, there exists a number of videos relying on both recommendation forces, which results in worse fitting results for BModel. In fact, only those warm videos are prone to be affected by both forces. BModel is better for most popular videos.

### V. Case Study: TV

As we have discussed, our dynamic popularity model can quantify the strengths of the recommendation forces through model fitting so that we can observe user behaviors and evaluate recommendation strategies. We conduct case study with TV episodes to illustrate this point.

#### A. Aggregating episodes

Both DModel and WModel assume that users make independent decision to view videos. In reality, some videos may be related such as the News videos about the same event and different episodes of the same TV series. Thus, it is reasonable and necessary to aggregate these related videos into a single composite video before we study the recommendation strategies.

There are a total number of 9010 TV episodes used for case study, after excluding 695 badly fit TV episodes with large NMSE. For each TV series, we aggregate all episodes and different versions (e.g., high/low definition versions, different language versions) into a composite video by averaging their parameters $\alpha$, $\beta$, etc. Each composite video is classified as a d-recommended video if most of its episodes achieve better fitting with DModel, or a W-recommended video otherwise. After aggregation, there are 190 d-recommended and 112 w-recommended composite videos. Most d-recommended videos are newly produced TV series while most w-recommended videos are old TV series which are uploaded to the system recently. They are investigated separately below because of their different recommendation strategies and user behavior patterns.

#### B. D-recommended videos

Fig. 3(a) shows the scatter plot of log($q$) v.s. log($\alpha$) of all the episodes from two sampled TV series before aggregation. The cross dots represent the parameters of a hot TV while the round dots represent a warm TV. Some selected dots are annotated with episode numbers. We can observe that the first episodes attract more views than the following episodes for both TV series. As episode number increases, attractiveness decreases gradually, indicating that the first episode is the most attractive one and not all users can keep up with the update of TV episodes.

Fig. 3(b) shows the scatter plot of log($q$) v.s. log($\alpha$), and Fig. 3(c) shows the scatter plot of $\frac{\text{view count}}{qN}$ v.s. $\alpha$ for the 190 composite videos. Both scatter plots show strong positive Pearson correlations. Note that users are likely to make independent decisions to view composite videos, and thus it is fair to compare recommendation resources allocated to composite videos. Fig. 3(b) indicates that a strong positive correlation exists between video intrinsic attractiveness and direct recommendation strength. Concentrating on recommending popular videos is a reasonable strategy given limited direct recommendation resource. Fig. 3(c) shows large $\alpha$ tends to achieve high completion fraction, confirming that video information will be diffused more sufficiently if it is allocated with more recommendation resource.

---

Footnotes:

1. The aforementioned procedure of removing cold videos are done after fixing total user population for each video type.
2. Due privacy issue, we don’t report the actual user number.
C. W-recommended videos

Similar to Fig. 3a, we investigate the relationship between intrinsic popularity $q$ and its WOM recommendation rate $\beta$ for episodes within two sampled TV series. Fig. 3d is the scatter plot of $\log(q)$ and $\log(\beta N)$ of all the episodes from a hot TV series with 40 episodes and a warm TV series with 39 episodes. An interesting observation is that both the first several episodes and the last several episodes are very popular. Such different user behaviors from Fig. 3a are due to the fact that episodes of d-recommended TV series are uploaded slowly, typically one or two episodes per day. However, all episodes of most w-recommended TV series are uploaded together because they are produced several months or years ago. Given all episodes available, users are prone to browse the first and the last several episodes.

For episodes in Fig. 3a and Fig. 3d, there is a trend that recommendation rate increases with episode number. We believe that earlier episodes act as “advertisements” of the later episodes, such that user reaction rate becomes faster and faster. An exception is the last several episodes of w-recommended videos. We believe that these videos are browsed just following the browsing of the first several videos.

Fig. 3c shows the scatter plot of $\log(q)$ v.s. $\log(\beta N)$ and Fig. 3f shows the scatter plot of $\frac{\text{view counts}}{qN}$ v.s. $\frac{1}{qN}$ for the 112 w-recommended composite videos. Surprisingly, a strong negative correlation appears between $\log(\beta N)$ and $\log q$ in Fig. 3c indicating that users prefer recommending less popular videos. This observation is consistent with our intuition. People like to share interesting but uncommon things with friends. Fig. 3f shows a strong positive correlation between $\frac{\text{view counts}}{qN}$ v.s. $\frac{1}{qN}$, implying that initial seed population is very essential for information diffusion.

VI. Related Work

Some previous studies have proposed several models of dynamic video popularity. In [3], the authors studied video sharing in online social network and proposed a probabilistic model that matches the observed dynamic popularity distribution. But their model does not depict individual videos’ popularity evolutions. Avramova et al. [5] proposed a closed-form expression for a video’s popularity evolution, which can be degenerated into either a power-law or exponential decay function. However, it is an ad-hoc model and does not reveal information diffusion process. In [6], although the authors consider both recommendation mechanisms, their model is rather complicated and cannot be easily applied to quantify the recommendation strength.

There exist works that studied popularity prediction. Szabo and Huberman [7] found that a linear relationship exists between a video’s early and future popularity, and thus future popularity can be predicted through multiplying early popularity by a linear coefficient learned from past datasets. In [3], the authors generalized the Szabo-Huberman method by considering multiple early popularity and video similarity. Ahmed et al. [4] classified videos by their popularity evolution patterns and predicted future popularity based on the classification. These predictive models focus on prediction accuracy but lack explainable and measurable factors, hence cannot give theoretical insights.

Video popularity is an important reference for resource allocation in content delivery. Researchers have studied various schemes by incorporating popularity dynamics. For example, Zhou et al. [2] proposed a mixed CDN caching strategy for both age-sensitive and popularity-stable videos. Hu et al. [10] leveraged community based social video viewing behaviors to
develop a cloud CDN content placement scheme. [11] tried to detect popular videos as CDN caching candidates by mining main steam media. However, in these works the popularity analysis are mainly based on measurements and theoretical models are not proposed.

VII. CONCLUSION

In this paper, we develop a dynamic model to depict video popularity evolution. Specifically, two recommendation mechanisms, direct and WOM recommendations, are incorporated in the model as the driving forces of video information diffusion process. Our model provides a systematic approach to quantifying video recommendation forces and other factors, and is useful for characterizing user behaviors and evaluating recommendation strategies, which are illustrated with a case study of TV episodes. Extending our model to involve more factors and predicting video popularity evolution will be our future work.

REFERENCES

[1] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, “Analyzing the video popularity characteristics of large-scale user generated content systems,” IEEE/ACM Transactions on Networking, vol. 17, no. 5, pp. 1357–1370, 2009.
[2] Y. Zhou, L. Chen, C. Yang, and D. M. Chiu, “Video popularity dynamics and its implication for replication,” IEEE Transactions on Multimedia, vol. 17, no. 8, pp. 1273–1285, 2015.
[3] H. Pinto, J. M. Almeida, and M. A. Gonçalves, “Using early view patterns to predict the popularity of Youtube videos,” in ACM WSDM, 2013.
[4] M. Ahmed, S. Spagna, F. Huici, and S. Niccolini, “A peek into the future: Predicting the evolution of popularity in user generated content,” in ACM WSDM, 2013.
[5] Z. Avramova, S. Wittervongel, H. Bruneel, and D. D. Vlaeschaert, “Analysis and modeling of video popularity evolution in various online video content systems: Power-law versus exponential decay,” in the First International Conference on Evolving Internet, 2009.
[6] J. Wu, Y. Zhou, D. M. Chiu, and Z. Zhu, “Modeling dynamics of online video popularity,” in IEEE IWOQS, 2015.
[7] G. Szabo and B. A. Huberman, “Predicting the popularity of online content,” Communications of the ACM, vol. 53, no. 8, pp. 80–88, 2010.
[8] H. Li, H. Wang, J. Liu, and K. Xu, “Video sharing in online social networks: Measurement and analysis,” in ACM NOSSDAV, 2012.
[9] Authors, “Appendix, supplied as additional material appendix.pdf.”
[10] H. Hu, Y. Wen, T.-S. Chua, Z. Wang, J. Huang, W. Zhu, and D. Wu, “Community based effective social video contents placement in cloud centric cdn network,” in IEEE ICME, 2014.
[11] A. Lobzhanidze and W. Zeng, “Proactive caching of online video by mining mainstream media,” in IEEE ICME, 2013.

APPENDIX

Derivation of Eq. 8

Let \( x' = p \) and thus we have \( x'' = p' \). Since

\[
p = \frac{dx}{dt} = \frac{dx}{dp} \cdot \frac{dp}{dt} = \frac{dx}{dp} \cdot p,
\]

we have \( p' = p \cdot \frac{dp}{dt} \).

Note that \( s = N - \frac{1}{q} \). Substituting it back to Eq. 7, we have

\[
x'' = x' (\beta qN - \beta x - 1)
\]

\[
\Rightarrow p \cdot \frac{dp}{dx} = p(\beta qN - \beta x - 1)
\]

\[
\Rightarrow \frac{dp}{dx} = \beta qN - \beta x - 1
\]

\[
\Rightarrow p = (\beta qN - 1)x - \frac{1}{2} \beta x^2 + C
\]

where \( C \) is a to-be-determined constant. With the initial condition \( x(0) = \beta q x_0 (N - \frac{1}{q}) \), we can solve for \( C \) and the two roots of \(-\frac{1}{2} \beta x^2 + (\beta q N - 1)x + C = 0\) are

\[
x_1 = qN + \frac{\varphi - 1}{\beta}
\]

\[
x_2 = qN + \frac{-\varphi - 1}{\beta}
\]

It is not difficult to show that \( x_2 < x_0 < x_1 \). Therefore,

\[
\frac{x - x_1}{x - x_2} = -g(t) \Rightarrow x(t) = \frac{x_1 + g(t)x_2}{1 + g(t)}
\]

With the initial condition \( x_0 \), we can get \( g(t) = \frac{x_1 - x_0}{x_0 - x_2} e^{-\varphi t} \).

Proof of Proposition 3 Let \( f(x_0, \beta) = x_1 = qN + \frac{\varphi - 1}{\beta} \), where \( \varphi = \sqrt{(\beta q N - 1)^2 - (\beta x_0 - 1)^2 + 1} \). Since

\[
\frac{\partial f}{\partial \beta} = \frac{(qN - x_0)(\beta qN + \beta x_0 - 1)}{\varphi}
\]

we have

\[
\frac{\partial f}{\partial \beta} = \frac{\beta (qN - x_0) + \varphi - 1}{\varphi^2} = \frac{\beta (qN - x_0) + \beta (x_1 - qN)}{\varphi} > 0
\]

Hence, final population \( x_1 \) always increases as \( \beta \) increases.

\[
\frac{\partial^2 x_1}{\partial \beta^2} = \frac{\partial x_1}{\partial \beta} = \frac{(x_1 - x_0)(\frac{\partial x_1}{\partial \beta} + \varphi)}{\varphi^2} = \frac{(x_1 - x_0) \left( 1 - \frac{\partial x_1}{\partial \beta} - \varphi \right)}{\varphi^2}
\]

\[
x > x_0. \text{ By letting } \left( 1 - \frac{\partial x_1}{\partial \beta} - \varphi \right) = 0, \text{ we derive } \beta = \frac{1}{2(qN + x_0)}.
\]