**On Celebrity, Epidemiology and the Internet**

|  |  |  |
| --- | --- | --- |
| Marily Nika | Gergana Ivanova | William J. Knottenbelt |
| Department of Computing | Department of Computing | Department of Computing |
| Imperial College London | Imperial College London | Imperial College London |
| London SW7 2AZ, UK  marily@imperial.ac.uk | London SW7 2AZ, UK gi12@imperial.ac.uk | London SW7 2AZ, UK wjk@doc.ic.ac.uk |

ABSTRACT

The proliferation of the internet has created new opportuni-ties to study the mechanisms behind the emergence and dy-namic behaviour of online popularity and celebrity. In this paper we examine how common epidemic models, specif-ically SIR and SEIR models, can be applied to model the evolution of outbreaks of celebrity interest on the internet. A major challenge when using such models is to parameterise them to fit data as an outbreak unfolds over time, with-out knowing the initial number of susceptibles in the target population. We present a methodology capable of fitting the model’s parameters from a single trace, while the outbreak unfolds, and of forecasting the epidemic’s progression in the coming days. We present results on three kinds of data: simulated epidemic data, data from a real Influenza virus outbreak and data from music artists BitTorrent download and YouTube video views activity.

Similar to a disease’s behaviour, an outbreak of celebrity interest starts with a few susceptible individuals who are exposed to an originating event and some of whom become“infected”. These individuals then interact with others, pass-ing on the disease or information. Eventually the infected individuals recover/lose interest and the outbreak dies out.

By way of example, let us consider the outbreak of pub-lic interest following the death of the music artist Whit-ney Houston. The left side of Figure 1 presents Whitney’s YouTube music videos’ views as recorded immediately af-ter her death. We observe a rapid surge in interest which remains particularly high for 7 days, and then gradually be-gins to drop down to previous levels. The right-hand side of Figure 1 presents occurrences of Influenza-like Illness in-cidents as reported in 2013 in Kansas [12]. The two curves appear to share a similar shape profile.

As noted in [15], until very recently, the study of celebrity was widely held in“serious”academic circles to be a marginal

|  |  |
| --- | --- |
| Keywords  Epidemiology, Mathematical Modelling, Celebrity | pursuit. However, in the last two decades, many disciplines, from sociologists to computer scientists, have begun to ac-tively study this ubiquitous modern status phenomenon. |

Despite numerous qualitative analyses of celebrity [15, 30,

|  |  |  |
| --- | --- | --- |
| 1. | INTRODUCTION | 4], there is very little quantitative understanding of the ori- |
| gin and evolution of celebrity. While some researchers have |

n. Ce-leb-ri-ty A person who has a prominent profile and commands a great degree of public fascination and   
influence in day-to-day media.

Celebrities pervade our social existence and media, not only as the faces of popular culture but also as the focus of intense public interest. In the information age, celebrity is born and spreads relatively quickly thanks to the rapid dissemination of information via multiple channels, many of them internet-based (e.g. social networks, video websites and peer-to-peer file sharing networks). For the same rea-son, outbreaks of celebrity tend to be very ephemeral as public interest is at first excited and then dissipates as focus shifts. Indeed, Andy Warhol’s famous prediction that “in the future, everyone will be world-famous for 15 minutes”[34] seems to becoming reality.

Permission to make digital or hard copies of all or part of this work for

|  |
| --- |
| �������������������������������������������������������������������������personal or classroom use is granted without fee provided that copies are  ���������������������������������������������������������������������not made or distributed for profit or commercial advantage and that copies  �����������������������������������������������������������������������bear this notice and the full citation on the first page. To copy otherwise, to  republish, to post on servers or to redistribute to lists, requires prior specific  ������������������������������������������������������������������������  permission and/or a fee.  ������������������������������������������������������������������������ValueTools’13, December 10 – 12 2013, Turin, Italy  ������������������������������������������������Copyright 2013 ACM 978-1-4503-2539-4/13/12 ...$15.00.  ���������������������������������������������� |

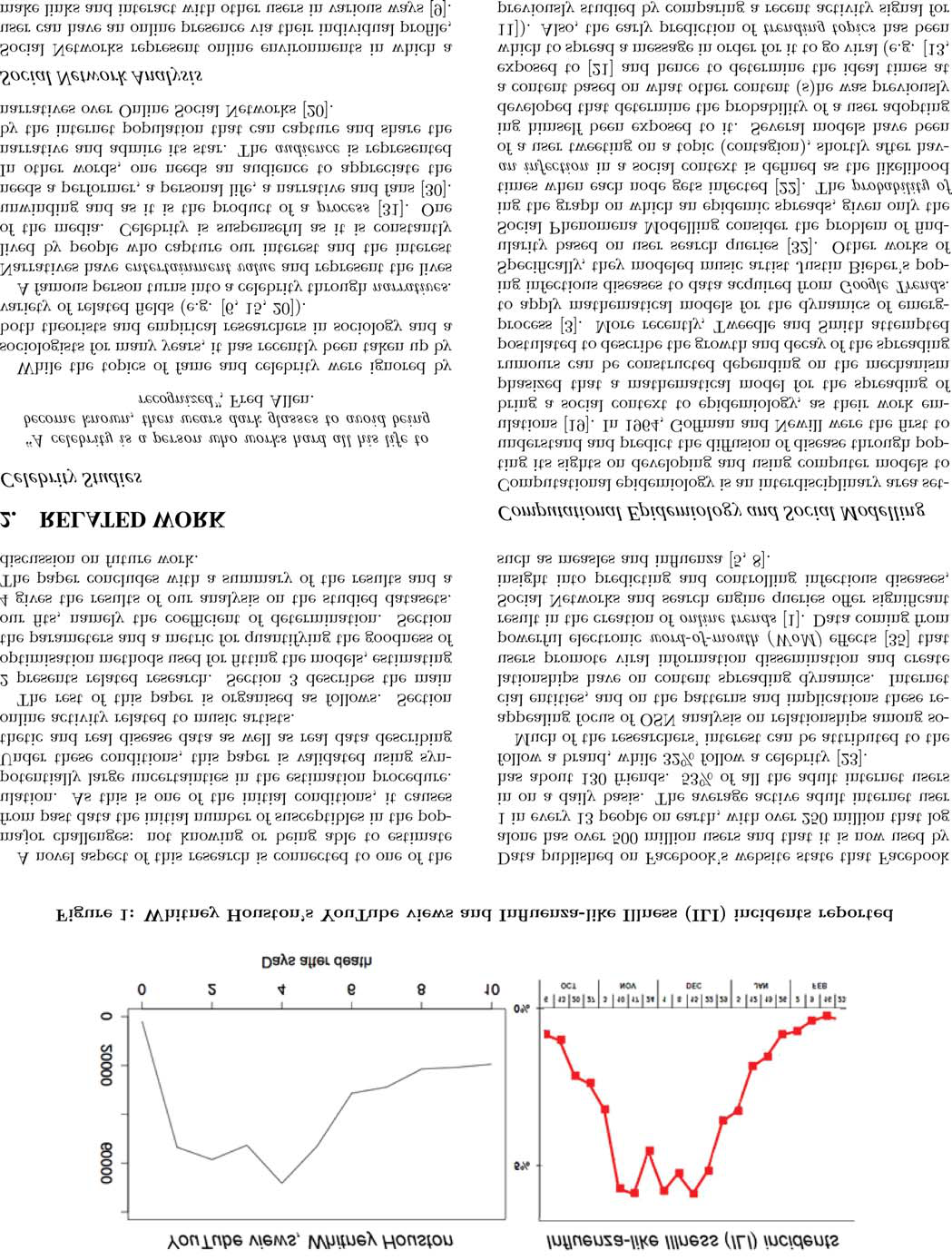
������������������������������������������������������������������������������

developed quantitative models of the popularity dynamics of certain items of online content such as Wikipedia articles [28] and YouTube video views [18], quantitative studies of celebrity are still at an early stage [10, 32].

By contrast, in the domain of disease modelling, there has been a large amount of work on epidemiology (e.g. [26, 3, 36, 25]). Our research aims to explore to what extent the lessons learnt in epidemiology can be used in a study of the way celebrity spreads on the Internet. It turns out there is no direct translation, because this kind of research comes with some extra challenges, such as not knowing the initial number of susceptibles in an online user population.

The main contribution of this paper is the study of how epidemiological models can adapted for modelling and pre-dicting the spread of celebrity. Two classical infectious dis-ease models, namely the SIR and SEIR models, are used within a model parameter fitting framework that takes as input a truncated dataset describe some “outbreak” of on-line activity following an event involving a celebrity. Given an outbreak, a prediction technique is developed for the on-line fitting of infectious disease model parameters using an optimization method that employs the Nelder–Mead algo-rithm with a least-squares-based objective function.

���



���

a topic to a large collection of historical activity signals for trending and non-trending topics [7], as well as the popular-ity life-cycle of YouTube videos which has been studied ei-ther by examining their popularity distribution versus their age [2] or by analyzing early measurements of view data [17].

3. METHODOLOGY

3.1 Modelling Epidemic Processes

We are developing optimisation-based frameworks based on traditional epidemiological models [33], in order to shed light on the following question: Given a snapshot of a social behaviour with some behaviour occurrences (i.e. an emerg-ing trend), how early on in the outbreak will we be able to predict aspects of its future evolution? We study the rela-tionship between popularity dynamics and virus infectivity by calculating certain time points of interest. As illustrated in Fig. 2, these are: the time when the epidemic reaches its peak in terms of number of infectious individuals, the time by which at least half of those individuals have recovered, and the time when the epidemic ends.

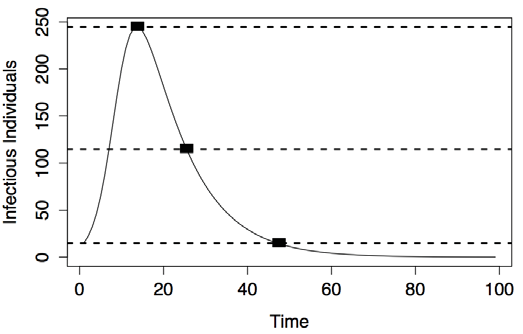


Figure 2: Sample infectious disease outbreak data with marked points of interest.

3.1.1 The SIR model

An epidemic is said to arise in a community when cases of a disease or other health-related events occur in excess of normal expectancy. We define an outbreak as an event in a celebrity’s career or personal life that has attracted the interest of the media, such as a TV appearance, a gig, a

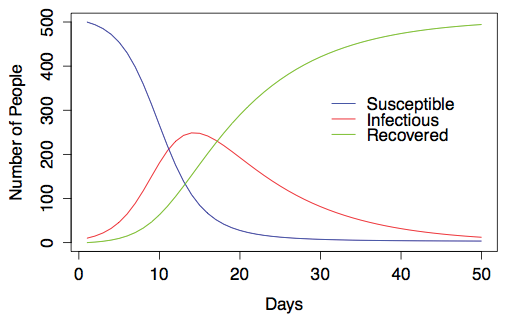


Figure 3: Sample run of the SIR model with pa-rameters β = 0.001, γ = 0.1 and initial conditions S0 = 500, I0 = 10 for 50 days.

• I(t) the number of individuals who are infected by the disease at time t with rate β,

• R(t) the number of individuals who have recovered from the disease at time t. We assume that the rate of recovery γ is constant and therefore the infectious period follows the exponential distribution.

The initial values of SIR need to satisfy the conditions:

S(0) = S0 > 0 (1)

I(0) = I0 > 0 (2)

R(0) = 0 (3)

To illustrate how the SIR model evolves, we solve the sys-tem of differential equations above for chosen input values: β = 0.001, γ = 0.1 with initial conditions S0 = 500, I0 = 10. Consider the resulting numbers of the susceptibles, infec-tious and recovered individuals through time in Fig. 3. Note how the equality N = S + I + R is preserved throughout.

3.1.2 The SEIR model

The main difference the SEIR model has compared to the SIR model, is an additional subpopulation, the Exposed E, consisting of individuals who are infected but not yet in-fectious. If we assume that the sojourn time of individuals in the latent period follows an exponential distribution with expectation α−1, the differential equations for the model are:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| release of a new single/album, or even larger events such as | dS(t) | | | =S()It) | () |
| a marriage, divorce, scandal or death. |
|  | dt | | | = −S()It) | () |
| Kermack and McKendrick’s classical models of 1927 have |
| suggested the use of Ordinary Differential Equations (ODEs) | dE(t) | = βS(t)I(t) − αE(t) | | | () |
|  |
| [14] as an appropriate modelling formalism. The most basic, | dt | () |
| the SIR model, counts the number of susceptible, infected, |
| and recovered individuals in a population. The SIR model | dI(t) | | = αE(t) − γI(t) | | () |
| and other derived infectious disease models (e.g. [25, 26]), | dt | | () |

allow us to answer questions such as how many people need to be vaccinated to prevent an epidemic? or how many people will be infected at a particular point in time? Given a closed population of individuals, we define three subpopulations:

• S(t) the number of individuals who are susceptible to become infected by the disease at time t,

|  |  |  |
| --- | --- | --- |
| dR(t) | = I() | () |
| dt | = I() | () |

Fig. 4 presents a sample evolution of the SEIR model with pre-supplied parameters. Compared to the SIR model’s evo-lution, the SEIR model’s curve is more platykurtic and its infectious peak is reached later in time.

���

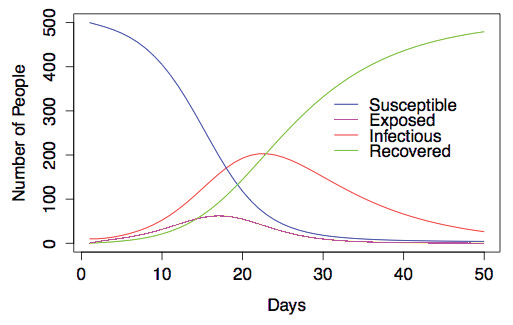


Figure 4: Sample run of the SEIR model with pa-rameters β = 0.001, α = 0.5, γ = 0.1 and initial condi-tions S0 = 500, E0 = 0, I0 = 10 for 50 days.

3.2 Model Fitting Procedure

An important application of mathematical models is to estimate parameters that cannot be measured directly. Here we discuss how we fit the parameters of our models in the context of ongoing outbreaks. We particularly consider the challenge of estimating the initial number of susceptibles in populations where this quantity is not known, and there is no principled way for estimating it. Traditional methods for estimating parameters in SIR/SEIR models involve only the estimation of β, γ and (where applicable) α. This is because the initial number of susceptibles has traditionally been considered to be a known quantity or one that can be readily estimated from the context [24, 29, 33].

3.2.1 Isolating Outbreaks

observation exceeds the mean of the observations so far by three standard deviations (cf. Fig. 5). We regard a particular outbreak as having ended when the standard deviation of a sliding window formed from the most recent k observations falls to or below the level observed just before the start of the outbreak.

3.2.2 Online Model Fitting

We attempt to make predictions while each outbreak un-folds, over time. For that reason, we apply our fitting method-ology on truncated datasets. For each dataset, we start by taking the first 3 observations of the outbreak. We then create a new truncated dataset by adding 1 more new ob-servation at a time, until the end of the outbreak.

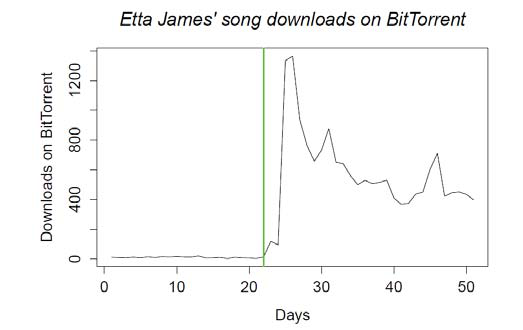
Parameters need to be estimated for each truncated dataset. The vector of parameters that needs to be estimated for SIR models is β, γ and S0 and for SEIR models β, γ, α and S0. For technical reasons to do with the optimisation method employed and the fact that all rates are known to be pos-itive, we actually work in log space and fit log(β), log(γ), log(α) (where applicable) and log(S0).

3.2.3 Searching the Parameter Space

In order to perform a search of the parameter space for the set of model parameters which gives the best least-squares fit to the data, we make use of the Nelder–Mead method. The Nelder–Mead algorithm is a method for multidimen-sional unconstrained optimization that does not require the calculation of derivatives. It is widely used to solve parame-ter estimation and maximum likelihood problems, where the objective function is not smooth [16].

In our case we make use of a least-squares-based objective function that characterises how well a candidate model fits the real data. That is, our approach produces a solution that minimizes the sum of squared residuals. Algebraically this corresponds to minimising

|  |  |  |  |
| --- | --- | --- | --- |
| Isolating an outbreak from background trend data re- | S = | �(yi − f(xi, θ))2 | (8) |
| quires rules which define the start and end of an outbreak. |



where yi is the observed value, and the model is f(xi, θ) where θ is the vector of unknown parameters. The model fits are performed by solving first-order ODEs using the R package lsoda. Note that it is important to specify a small number for the absolute error tolerance, which determines the error control performed by the solver. Alternatively, one can specify the maximum value for the integration step-size.

Regarding initial conditions, we take I0 to be the number of infectious individuals on the first day of the outbreak, while R0 is assumed to be 0. In the case of the SEIR model, we also assume E0 to be 0.

To mitigate the likelihood of the Nelder–Mead optimisa-tion procedure becoming trapped in a local minimum, we restart it with 20 different random initial parameter vectors (sensibly constrained such that γ > β > 0 for example), and select as our final candidate that vector which yields the lowest S across all runs.

|  |  |  |  |
| --- | --- | --- | --- |
| Figure 5: | Outbreak detection in action (vertical | 3.2.4 | Assessing Goodness of Fit |
| line) on downloads of Etta James’ songs. | |

In order to assess how well a chosen parameter vector fits

While it is often obvious in retrospect to link the start of an outbreak to some particular activity or event, such a link may not be obvious at the time, and/or may not always

a truncated dataset, we make use of the coefficient of deter-mination, denoted as R2. Normally reported in the context of techniques such as regression, R2describes the proportion

be present. For the purposes of this paper, we deem an of the total variation present in the observations explained

observation to mark the beginning of an outbreak if the next by the model. Assuming that yi are the observed data points

���

and fi are the model predictions, the mean of the observed data is given by ¯y = (�n i=1yi)/n . Then we calculate the to-

tal sum of squares, SStot, which is proportional to the sam-

FluView Web Portal1, we obtained a dataset of influenza positive tests (summed over all subtypes of the flu virus) reported to the CDC for the 2012/2013 Influenza season.

|  |  |  |
| --- | --- | --- |
| ple variance, and the residual sum of squares SSres, which | 3.3.3 | MusicMetric Data |
| gives a measure of how far the estimated values are from the |

observed. The formulae are the following:

We were able to gather time-series data for BitTorrent downloads and YouTube video views of various music artists

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SStot = | n �(yi − ¯y)2 | (9) | | using the MusicMetric API. This is an online artist analytics toolbox that contains detailed information on fan trends and popularity for particular artists. | |
| i=1 | | | |
| SSres = | n �(yi − fi)2 | | (10) | 4. | RESULTS |
| i=1  Then the coefficient of determination is given by | | | | In this section, we present results illustrating the applica-tion of our online model fitting methodology to our different datasets. We use the coefficient of determination as a metric to assess the efficacy of our models. | |
| R2= 1 −SSres  SStot | | (11) | |
| 4.1 | Synthetic datasets |
| Normally, the value of R2will be in the range between 0 and 1. The closer R2is to 0 the least improvement our model | | | |
| SIR Data | |

has made over the simple model of taking the average of the observed data as our fitted value. The closer R2is to 1 the better our model explains the variability in the data. As can be observed by the formula above, if SSres > SStot, then R2can have negative values as well. In such situations, the mean of the data provides a better estimate than the model fitted values, thus meaning the model should be discarded.

The artificial dataset used in this section is shown in Fig. 6 and is generated from the SIR model with parameters β = 0.001, γ = 0.1 and initial conditions S0 = 500, I0 = 10. At a very early stage, and operating on a single set of only 8 observations, our model manages to predict with surprising precision not only that in 4 days there will be a peak of infectiousness, but also the number of infectious individuals

|  |  |  |  |
| --- | --- | --- | --- |
| 3.2.5 | Confidence Intervals on Model Trajectories | at that point. As time progresses, our fit becomes more and more stable and adjusts only slightly with the addition of | |
| The evolution of any realised trajectory of an epidemic | | new observations. | Finally, we can see that the estimated |

process is stochastic in nature. We therefore use multiple independent runs of Gillespie’s Stochastic Simulation algo-rithm [27] in order to capture the possible variation in the

best fit parameters are very close to their true values, the curve fits the data points well and the confidence intervals are providing a good indication of the predicted values.

|  |  |
| --- | --- |
| number of infected individuals observed at every time step given our best-guess model parameterisation. | SEIR Data |

Specifically, for the set of simulation generated observa-tions at each time point t and a confidence level of (100−c)%, we report the lower end point of the confidence interval as the cth percentile of the observations and the upper end point of the confidence interval as the (100 − c)th percentile of the observations.

Naturally, this formulation does not take into account the additional uncertainty that may be associated with the model parameterization itself. We acknowledge that this is-sue is important and needs to be considered in future work.

Similarly, we generate synthetic data from the SEIR model as shown in Fig. 7, with parameters β = 0.001, α = 0.5, γ = 0.1 and initial conditions S0 = 500, E0 = 0, I0 = 10. We manage to predict the curve and the peak before it actu-ally occurs with a good precision. Note that because of the extra parameter we initially observe that the curve is much smoother and changes much more with additional ob-servations. However, after observing 25 data points, the fit manages to predict the tail well.

|  |  |  |  |
| --- | --- | --- | --- |
| 3.3 | Data Sources | 4.2 | Actual Influenza outbreak dataset |

3.3.1 Synthetic SIR/SEIR Data

Synthetic datasets generated by SIR and SEIR models with known parameters were generated using stochastic sim-ulation. A number of packages are suitable for this purpose including R, Dizzy and Matlab. The purpose of using syn-thetic datasets is to evaluate the ability of our methodology to recover model parameters using a single trace for which the ground truth is known.

3.3.2 Real Influenza Data

Influenza is one of the most common infectious diseases in humans, with regular annual outbreaks. One institution that reports on the impact of flu in the US is the Center for Disease Control and Prevention (CDC). From the CDC’s

SIR Data

This data set is taken from reports of the US Center for Disease Control (CDC) for the 2012-2013 Influenza season and provides the number of individuals testing positive for flu over time. As seen in Fig. 8, we manage to predict the peak in infectious individuals from only 7 observations to be around day 10 and of magnitude around 6 800. In reality, it occurs to be only 1 day later, with slightly more peo-ple infected, about 7 000. The accuracy of predicting from partial information on a single trace the time of the peak, the magnitude of the peak and the tail of the infection is remarkable.

1http://gis.cdc.gov/grasp/fluview/ fluportaldashboard.html

���

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 250 | *Days after outbreak = 8* | | | | | | | 350 | *Days after outbreak = 14* | | | | | | 350 | *Days after outbreak = 21* | | | | | |
| *Infectious Individuals = 136* | | | | | | | *Infectious Individuals = 260* | | | | | | *Infectious Individuals = 179* | | | | | |
| ooooo | | | | | | | |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  | | | | | | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| 4.3 | Case Studies of Music Artists | [2] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and |

S. Moon. I tube, you tube, everybody tubes: analyzing

Whitney Houston’s death, SIR model of YouTube views

Fig. 9 is based on an SIR model fit to YouTube video plays of Whitney Houston’s songs online immediately after her death on 11 February 2012. Note the huge jump in views on the day after the event, where views skyrocket from around 2 000 to 53 000 in only a day. We speculate that this effect is due to the intense social media activity and saturation news coverage surrounding the event. In fact our model does not manage to predict the peak before it occurs, as it is very early on on the outbreak. Also, while there is reasonable qualitative agreement between the fitted model and the data overall, the limitations of our current strategy for generating confidence intervals without due regard for parameter uncertainty become very apparent.

Whitney data, SEIR model of BitTorrent Downloads

Fig. 10 presents a SEIR model fit to the daily BitTorrent downloads of Whitney Houston’s music shortly after her death. The extra parameter allows for good flexibility in the model fit. Indeed, the fitted curve follows the data points fairly closely from day 14 of the outbreak. The fit remains relatively stable with the addition of new observa-tions, which allows us to predict the tail of the outbreak with a good amount of certainty.

Etta James SEIR BitTorrent downloads after her death

Turning now to an SEIR model of the BitTorrent downloads following the death of soul and blues singer Etta James on 20 January 2012, we observe in Fig. 11 that from day 5 of the outbreak the model is able to accurately predict the landing point of the downloading epidemic.

5. CONCLUSIONS

This paper represents a preliminary attempt to under-stand the origins and dynamic evolution of celebrity on the internet by drawing on, and extending, the classical theory of the epidemiological modelling of infectious diseases. It is promising that the proposed framework appears to be able to successfully recover the parameters of synthetic datasets at an early stage, and is flexible enough to be applied with some success to real data ranging from BitTorrent music download traffic and YouTube video views to Influenza in-cidence.

This effort forms part of a broader framework which aims to be able to answer questions such as: What sort of ac-tions create the greatest outbreaks of public interest? How long will a given increase of public interest last? and At what point in time the public interest will reach a peak?. Further, by applying quantitative models to the domain of modern music, we want to shed light on how the internet affects our preferences and evolving tastes in music artists. Wider application areas are also worthy of investigation, e.g. prediction of computer virus spread and mobile application downloads.

the world’s largest user generated content video system. In IMC ’07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, pages 1–14, New York, NY, USA, 2007. ACM.

[3] D. J. Daley and D. G. Kendall. Epidemics and rumours. Nature, 204(4963):1118–1118, 12 1964.

[4] David Marshall. Celebrity and Power: Fame in   
Contemporary Culture. University of Minnesota Press, 1997.

[5] A. Dugas, M. Jalalpour, Y. Gel, S. Levin, F. Torcaso, T. Igusa, and R. Rothman. Influenza forecasting with Google flu trends. Online Journal of Public Health Informatics, 5(1), 2013.

[6] K. O. Ferris. The Sociology of Celebrity. Sociology Compass, 1(1):371–384, 2007.

[7] D. Gao, W. Li, and R. Zhang. Sequential   
summarization: A new application for timely updated twitter trending topics. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 567–571, Sofia, Bulgaria, August 2013. Association for   
Computational Linguistics.

[8] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant. Detecting influenza epidemics using search engine query data. Nature, 457(7232):1012–1014, Nov. 2008.

[9] R. Gross and A. Acquisti. Information revelation and privacy in online social networks (the Facebook case).

In Proceedings of the 2005 ACM workshop on Privacy in the electronic society, pages 71–80.

[10] W. Hartmann, P. Manchanda, H. Nair, M. Bothner, P. Dodds, D. Godes, K. Hosanagar, and C. Tucker.

Modeling social interactions: Identification, empirical methods and policy implications. Marketing Letters, 19(3):287–304, December 2008.

[11] H.-W. Hu and S.-Y. Lee. Study on influence diffusion in social network. International Journal of Computer Science and Electronics Engineering (IJCSEE), 1, 2013.

[12] Influenza Division, Kansas Department of Health and Environment. Weekly influenza surveillance report, September 2013.

[13] A. Karnik, A. Saroop, and V. Borkar. On the diffusion of messages in on-line social networks. Performance Evaluation, 70(4):271 – 285, 2013.

[14] W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics–i. 1927. Bull Math Biol, 53(1-2):33–55, 1991.

[15] Kerry O. Ferris and Scott R. Harris. Stargazing: Celebrity, Fame, Social Interaction. Contemporary Sociological Perspectives. Taylor and Francis Ltd, 2011.

[16] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E.

Wright. Convergence properties of the nelder-mead simplex method in low dimensions. SIAM Journal of

|  |  |  |
| --- | --- | --- |
| 6. | REFERENCES | Optimization, 9:112–147, 1998. |
| [17] S. Leonardi, A. Panconesi, P. Ferragina, and |

[1] Y. Altshuler, W. Pan, and A. Pentland. Trends prediction using social diffusion models. CoRR, abs/1111.4650, 2011.

���

[18] H. Li, H. Wang, J. Liu, and K. Xu. Video sharing in online social networks: measurement and analysis. In Proceedings of the 22nd international workshop on Network and Operating System Support for Digital Audio and Video, NOSSDAV ’12, pages 83–88, New York, NY, USA, 2012. ACM.

[19] M. Marathe and A. K. S. Vullikanti. Computational

86:3200–3203, Apr 2001.

[27] M. Pineda-Krch. GillespieSSA: Implementing the Gillespie Stochastic Simulation Algorithm in R.

Journal of Statistical Software, 25(12):1–18, Feb. 2008. [28] J. Ratkiewicz, F. Menczer, S. Fortunato, A. Flammini, and A. Vespignani. Characterizing and modeling the dynamics of online popularity. CoRR, abs/1005.2704,

epidemiology. Commun. ACM, 56(7):88–96, July 2013. 2010.

[20] A. Marwick and D. Boyd. To See and Be Seen: Celebrity Practice on Twitter. Convergence: The International Journal of Research into New Media Technologies, 17(2):139–158, May 2011.

[21] S. Myers and J. Leskovec. Clash of the Contagions: Cooperation and Competition in Information   
Diffusion. In Proceedings of ICDM, 2012.

[22] P. Netrapalli and S. Sanghavi. Learning the graph of epidemic cascades. In Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE joint international conference on Measurement and Modeling of   
Computer Systems, SIGMETRICS ’12, pages 211–222, New York, NY, USA, 2012. ACM.

[23] Nielsen. State of the media: The social media report. Technical report, Q3 2011.

[24] B. Ottar, F. Barbel, and G. Bryan. Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. Ecological   
Monographs, pages 169–184, 2002.

[25] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks.

Phys. Rev. E, 63:066117, May 2001.

[26] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. Phys. Rev. Lett.,

[29] TL Burr, G Chowell. Observation and model error effects on parameter estimates in   
susceptible-infected-recovered epidemic model. Far East Journal of Theoretical Statistics, 2006.

[30] G. Turner. Approaching Celebrity Studies. Celebrity Studies, 1(1):11–20, Mar. 2010.

[31] J. Turner, M. Begon, and R. G. Bowers. Modelling pathogen transmission: the interrelationship between local and global approaches. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270(1510):105–112, 2003.

[32] V. Tweedle and R. J. Smith. A mathematical model of Bieber Fever: The most infectious disease of our time. Understanding the dynamics. March 2012.

[33] E. Vynnycky and R. White. An Introduction to Infectious Disease Modelling. 2010.

[34] Wikipedia. 15 minutes of fame. http:   
 //en.wikipedia.org/wiki/15\_minutes\_of\_fame.

[35] D. D. Y. Liu and R. Burnkant. Provide consumers with what they want on word of mouth forums.

iBusiness, 5(1A):58–66, 2013.

[36] D. Zanette and S. Risau-Gusm´an. Infection spreading in a population with evolving contacts. Journal of Biological Physics, 34(1-2):135–148, 2008.

���