Applied Mathematical Modelling 35 (2011) 5442–5447

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

An epidemiological approach to model the viral propagation of memes

Lin Wang *, Brendan C. Wood

University of New Brunswick Fredericton, NB, Canada E3B 5A3

ARTICLE INFO

Article history:
Received 13 April 2010
Received in revised form 22 April 2011
Accepted 26 April 2011
Available online 8 May 2011

Keywords:
Meme
Stability
Epidemic model
Infection

ABSTRACT

An epidemiological approach is adopted to develop a model of viral meme propagation. The successful implementation in the modelling of meme spread as reflected in Internet search data shows that memes may be treated as infectious entities when modelling their propagation over time and across societies.

1. Introduction and background

In this paper we study the viral propagation of memes. As defined by Dawkins [1], a meme is any cultural entity that appears to exhibit self-replication. In particular, this concept applies to human thoughts and ideas, which can be transmitted from person to person via social interaction. Memes are subject to a form of natural selection, which allows them to change over time and even evolve to become more effective replicators. This corresponds to people having different interpretations or incomplete understanding of existing ideas, which can eventually give rise to wholly new ideas. In this sense, memes are the cultural analogue of genes.

Memetic modelling techniques allow for estimation of public interests and predictions of what the public may find interesting in the future. They can model the acceptance of new ideas, the decline of old ones, and the rapid spikes produced by sensational events that happen around the world. Since the Internet plays such a central role in today's world and mirrors society in so many ways, Internet data can be used for tracking and verification of memetic spread, as investigated in [2].

In the age of the Internet, memes have the ability to spread around the world in very short periods of time. Since memes have the ability to spread amongst populations, they have an infectious quality [3,4]. They spread via social interaction, much the same way as infectious diseases. Depending on the type of meme, people may lose interest over time, so memes can die out as well. Viral memes, in particular, exhibit a characteristic spike early in the infectious stage, followed by a gradual decay of the number of infected people. This behavior is very similar to the dynamics of the classical SIR model [5].

The dynamics of memetic spread are somewhat similar to that of infectious disease, but with a few key differences. One key difference is that in the viral propagation of memes, people who have lost interest in a meme may become interested in the meme again through social interactions. In this work, we take this aspect into consideration, and employ the idea of compartmental modeling in infectious diseases to investigate the viral propagation of memes. We point out that the epidemiological approaches are also used to model the spread of ideas [6], rumors [7] and ideology [8].

* Corresponding author. Tel.: +1 506 453 4780; fax: +1 506 453 4705.
E-mail address: lwang2@unb.ca (L. Wang).

0307-904X/$ - see front matter
doi:10.1016/j.apm.2011.04.035

The rest of the paper is organized as follows. The model is formulated and analyzed in Section 2. In Section 3, historic Internet search data are used to verify the model according to real-world trends. The conclusion is given in Section 4.

2. Model formulation and analysis

The standard SIR compartmental modeling approach divides a given population into three disjoint classes of individuals that are susceptible, infective, and recovered, with class sizes denoted by $S(t)$, $I(t)$ and $R(t)$, respectively. The model presented here will be referred to as a viral memetic model, in order to pay respects to its heritage and to distinguish it from the classical SIR model.

Since a meme does not correspond to any known physical entity, some terms will be redefined such that they more accurately represent reality. With respect to viral memetics, the word *susceptible* is used to describe people who have not yet been exposed to a particular meme. *Infected* refers to people who take an active interest in the idea or concept that a meme represented, and therefore have a tendency to talk about the meme in social interactions. The term *recovered* applies people who have experienced the meme, but have no interest or have lost interest at some point. Infected individuals decay into recovered individuals at a rate proportional to the number of infected individuals. Recovered individuals do not spread the meme in question, but they are assumed to be still susceptible to becoming reinfected by the meme, though possibly at a lower success rate than an individual not yet infected. In effect, the infected population has a socially attractive effect on both the susceptible and recovered populations. We assume that the infection rate is proportional to SI, the reinfection rate is proportional to IR, and the recovery rate is proportional to I. Then our viral memetic model is defined as follows:

$$\begin{aligned} \frac{dS}{dt} &= -\alpha SI, \\ \frac{dI}{dt} &= \alpha SI + \beta IR - \gamma I, \\ \frac{dR}{dt} &= -\beta IR + \gamma I. \end{aligned} \tag{1}$$

Here parameter α is defined as the transmission rate proportionality constant, which represents the fraction of interactions which result in a susceptible individual becoming infected with a meme. Parameter β ($<\alpha$) is the reinfection rate proportionality constant, which controls the rate at which recovered individuals become reinfected by the meme. Finally, parameter γ is the proportionality constant of loss of interest, which governs the rate at which infected individuals lose interest in the meme and move into the recovered population. Clearly, if parameter β were to equal zero, the viral memetic model would collapse to become identical to the classical SIR model [5].

There are two additional points to note about this model formulation. First, is that $\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$, which means that population size is constant. While this certainly is not true with respect to the Internet, it is approximately true over short periods of time. This assumption also allows the system to be analyzed more easily. We should point out that this assumption is not crucial and a model with varying population size N can be dealt with similarly without any fundamental difficulty, but as a start we stick to the constant population here in this paper. Second is that the social terms, SI and IR, imply a fully connected network. That is, for the social interaction between the susceptible and infected populations, every individual in I has an equal effect on every individual in S and *vice versa*. In a real society, individuals are distributed by location and cluster in communities, such that two individuals within a community may have strong influence on one another, whereas individuals on opposite sides of a country likely have zero direct influence on one another. This model does not account for localized sub-populations, and assumes that all individuals have an equal influence on all others.

For any given initial condition is $S(0) = S_0 > 0$, $I(0) = I_0 > 0$, $R(0) = 0$, there is a unique solution to system (1) satisfying $S(t) \geqslant 0$, $I(t) \geqslant 0$ and $R(t) \geqslant 0$ for all $t > 0$.

Note that the total population size N is a constant, this population constancy condition can be used to decouple the recovery dimension R and reduce the system to a two-dimensional coupled system of S and I. This results in the following system:

$$\begin{aligned} \frac{dS}{dt} &= -\alpha SI, \\ \frac{dI}{dt} &= (\alpha - \beta)SI + (\beta N - \gamma - \beta I)I. \end{aligned} \tag{2}$$

From the second equation of (1),

$$\left.\frac{dI}{dt}\right|_{t=0} = (\alpha S_0 - \gamma)I_0 \quad \begin{cases} > 0 & \text{if} \quad \alpha S_0 > \gamma, \\ < 0 & \text{if} \quad \alpha S_0 < \gamma. \end{cases}$$

This implies if $\alpha S_0 > \gamma$, then $I(t)$ initially increases. Moreover, given $\alpha \geqslant \beta$, from (2), $dS/dt \leqslant 0$, we have $S \leqslant S_0$ and

$$\frac{dI}{dt} = I((\alpha - \beta)S + \beta N - \gamma - \beta I) \leqslant I((\alpha - \beta)S_0 + \beta N - \gamma - \beta I) := h(I).$$

Note that $h(I)$ is a quadratic function of I and $I(t)$ remains nonnegative for all $t \geqslant 0$, it then follows $\frac{dI}{dt} \leqslant h(I) \leqslant 0$ for all $t \geqslant 0$ if $(\alpha - \beta)S_0 + \beta N - \gamma \leqslant 0$. This implies that if $(\alpha - \beta)S_0 + \beta N - \gamma \leqslant 0$, then $I_0 > I(t) \to 0$ as $t \to \infty$ and so the meme dies out.

Clearly if $\beta N \leqslant \gamma$, then (1) has non-isolated equilibria in the form of $(\bar{S}, 0)$ with $0 \leqslant \bar{S} < S_0$. On the other hand if $\beta N > \gamma$, then (1) has an additional equilibrium $E = (0, I^*)$ with $I^* = N - \frac{\gamma}{\beta}$. Linearizing the system about E to obtain the Jacobian

$$\mathbf{J}_E = \begin{bmatrix} -\alpha\left(N - \frac{\gamma}{\beta}\right) & 0 \\ (\alpha - \beta)\left(N - \frac{\gamma}{\beta}\right) & -\beta\left(N - \frac{\gamma}{\beta}\right) \end{bmatrix}.$$

Clearly both eigenvalues of $\mathbf{J}_E$ are negative and thus E is locally asymptotically stable. Since there is no interior equilibrium with both components positive, by index theory [9], the equilibrium E is globally asymptotically stable as well.

The above analysis is summarized in the following theorem.

Theorem 1. *Consider system (2) with any given initial condition $S_0 > 0$, $I_0 = N - S_0 > 0$. If $\alpha S_0 > \gamma$, then $I(t)$ initially increases and whether the meme persists or dies out eventually depends on the value of $R_0 = \frac{\beta N}{\gamma}$. If $R_0 \leqslant 1$, then the solution $(S(t), I(t))$ to system (2) satisfies $(S(t), I(t)) \to (\bar{S}, 0)$ as $t \to \infty$, where $\bar{S} = \bar{S}(S_0, I_0) \in [0, S_0)$ is initial condition dependent, and the meme dies out. Moreover, if $(\alpha - \beta)S_0 + \beta N - \gamma \leqslant 0$, then $I_0 > I(t) \to 0$ as $t \to \infty$ and so the meme dies out monotonically. If $R_0 > 1$, then*

$$\lim_{t\to\infty}(S(t), I(t)) = \left(0, N - \frac{\gamma}{\beta}\right),$$

and the meme persists.

3. Validation and discussion

In former times, it would have been difficult to gauge memetic penetration amongst a population without conducting a large statistical survey and repeating it every so often. Thanks to the Internet, however, it is possible to use historical search engine data to view what people have been searching for. If one assumes that people perform Internet searches for items that happen to be on their mind, then the frequency of searches for specific strings is proportional to the aggregate penetration of a meme into the social psyche of a population.

The Google Trends tool provides access to Google's search engine data since 2004, wherein one may enter a search string and view a chart of the search frequency or *search volume index* for that search string. The search volume index does not represent the absolute number of searches over a give time span. Rather, Google normalizes the search data against a number of variables to allow underlying characteristics of the data sets to be compared [10]. Additionally, the data are subject to either fixed or relative scaling, so it is not directly possible to access the raw number of searches, which would be the most obvious way of validating the viral memetic model. To produce the data, Google performs statistical analyses on subsets of its search logs, such that resulting data have low margins of error. The source data of the charts are available for download as `.csv` files, with data points spaced at weekly intervals.

Then comes the problem of selecting appropriate search strings. Since this paper describes a viral memetic model, the memes investigated should be relatively new concepts, which did not reach their peak of popularity before 2004. Not all viral memes produce smooth curves of interest, since the spread of memes depends greatly on publicity received from periodic sources such as on-line news sites and popular bloggers, as well as social networks. The viral memetic model best handles memes featuring one large, smooth spike shortly following their birth, and then a gradual taper.

The least-squares fitting technique has been applied to several data sets (acquired from Google Trends on October 16, 2009), of which three are presented below with estimated parameter values given in the following Table 1.

The *Ultimate Showdown of Ultimate Destiny* is a meme which originated in 2005 surrounding the release of a humorous Flash-based cartoon video. It became an immediate hit, with views and searches spiking very early in its history. Fig. 1 presents the fitted infection curve $I(t)$ plotted alongside the Google Trends data. The exact search string to acquire the data was "ultimate showdown".

The popularity of an image provides the second example. An owl with an amusing caption containing Internet slang was also created in 2005, which exhibited a spike very similar to that seen in Fig. 1. The popularity of the *O RLY owl* never reached the same high as that of the Ultimate Showdown, but its infection curve exhibits similar properties. A plot of the meme's activity can be seen in Fig. 2. The exact search string was "o rly".

Contrasting to both the O RLY owl and the Ultimate Showdown memes is the concept of a *blog*. While the idea of a blog existed prior to 2004, it's presence in the social psyche underwent a massive boom in 2004–2008. This is a more long-term trend which appears to have reached its peak but will likely maintain a significant presence on the Internet due to its high inertia. Its plot can be seen in Fig. 3. The exact search string was "blog".

Table 1
Estimated parameter values.

Parameter	Showdown	O RLY	Blog
α	5.78e−1	3.39e−3	1.62e−4
β	3.91e−4	3.35e−3	1.52e−4
γ	1.26e−2	3.35e0	3.00e−2
N	3.00e1	1.00e3	2.00e2

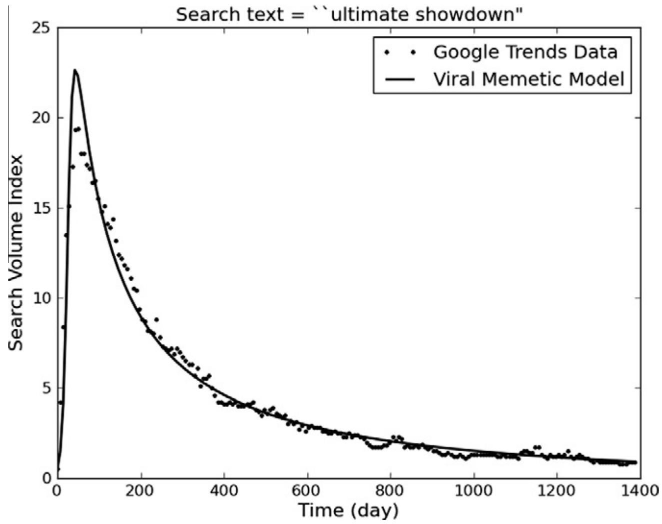

Fig. 1. Parameter optimization on "ultimate showdown" from December 2005 to October 2009.

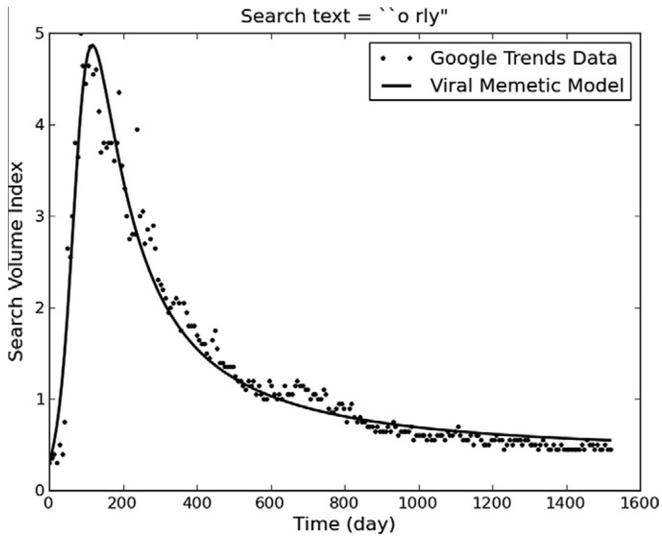

Fig. 2. Parameter optimization on "o rly" from August 2005 to October 2009.

Notice that parameter α is significantly smaller for the blog meme than the corresponding α parameters for the Ultimate Showdown and O RLY memes, both of which spike immediately and "burn out" shortly thereafter. The blog's small α parameter accounts for its slowness in growth. The Ultimate Showdown and O RLY memes both exhibit sharp initial spikes, so it is no surprise that their α parameters are so close in value. Parameter N represents the constant population size, which, due to Google's data normalization and scaling, has a convoluted meaning. N is best interpreted as a maximum search volume index, though the search volume index has no theoretical maximum. For the parameter estimation, N has been chosen to be an arbitrarily large value which is held constant throughout the least squares fitting. Ideally, N should be chosen such that it is quite large compared to the search volume index in the data. Both the O RLY and blog memes have produced good fits for large values of N, while the Ultimate Showdown meme required a smaller N in order to find good parameter estimations.

Based on these optimized parameters, the model can provide a prediction as to at what level the infected population will stabilize, if at all. In the case of blogs, it appears that the Google search index has peaked will stabilize at approximately 2.6. Note that this assumes a constant population, which is not necessarily true for such a long-term and inertial trend like blogs. Both the Ultimate Showdown and O RLY memes are expected to die out.

The parameter sets computed for Figs. 1 and 2 have produced roughly similar curves, although the parameter values differ significantly. For the test cases analyzed in this paper, the parameters α, β, and γ all have very small values, due to the large time scales of the data sets and the relatively large total population N necessary to support the model.

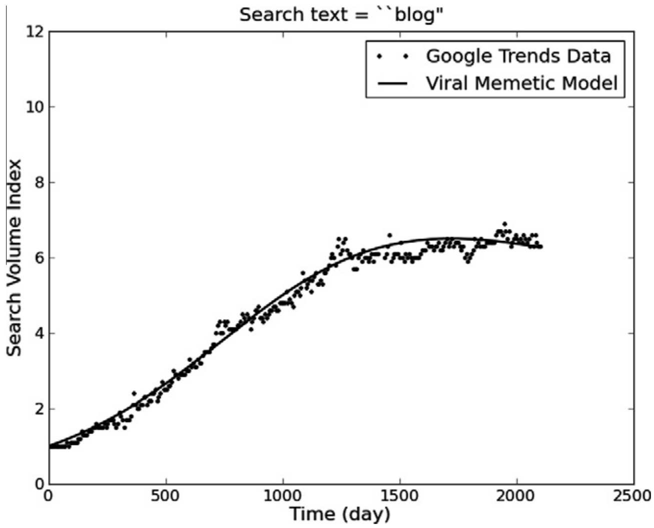

Fig. 3. Parameter optimization on "blog" from January 2004 to October 2009.

While *N* has been referred to as the population parameter, in these cases, *N* represents the total search volume against which the search volume index is computed. While performing the least-squares fitting of parameters, it became apparent that the value of *N* is largely arbitrary, as long as it was greater than the peak of the memetic trend. More work should be undertaken to quantify and understand the role of *N* in these models. The assumption of a constant *N* is reasonable over short periods of time but is not truly valid for the Internet or society as a whole. Indeed, the number of Internet users worldwide has roughly doubled since 2004.

The three previous data sets are smooth enough to allow close fitting of the viral memetic model, which is an indicator of the model's validity for at least some viral spreading of memes. However, not all data sets are smooth with a single peak. Take, for example, the net neutrality meme. Net neutrality is subject to an on-going debate in the United States of America as of the time of this writing, and has been under discussion since 2006. The debate is a slow process, which only receives major press when there are significant new developments. Press releases and media reports are major drivers of memes in a society, and as a result can cause successive minor epidemics. Fig. 4 shows the result of least squares parameter estimation for the entire history of the net neutrality debate. After hitting the first peak reasonably well, the model levels out and does not provide a good approximation of either subsequent peaks or the low base rate of searches. In order to improve results, it is necessary to reduce the size of the data set so that it only includes one major event. For example, reducing the data to only the first 85 weeks produces the fit seen in Fig. 5, which provides a higher degree of accuracy.

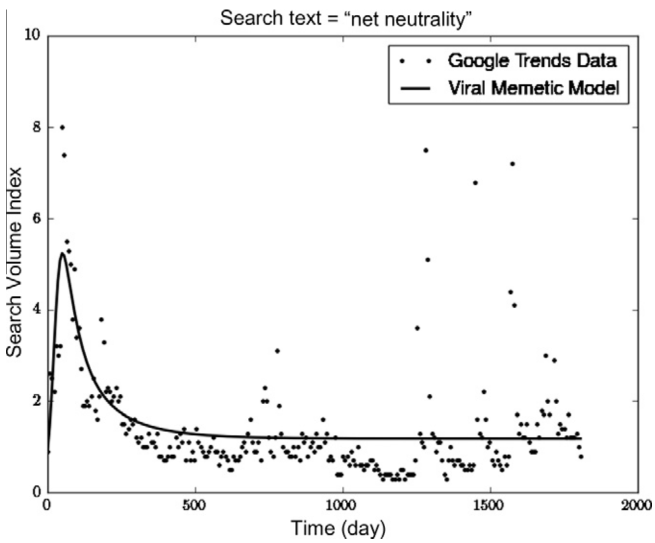

Fig. 4. Parameter optimization on "net neutrality" from April 2006 to April 2011.

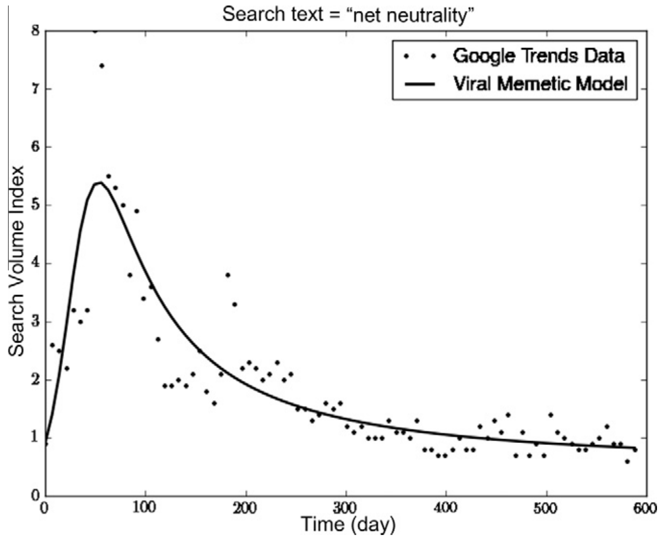

Fig. 5. Parameter optimization on "net neutrality" from April 2006 to December 2007.

The inability of the model to handle multiple peaks is a result of some of the simplifying assumptions made in constructing the model. Besides the constant *N*, the infection and decay parameters α, β and γ are treated as independent constants. In the real world, memes do not exist in isolation and their infection and decay parameters would be functions of other memes already present. This makes the parameter values time-dependent, which is not accounted for in this model. To properly model the infection and decay parameters, the entire memetic ecosystem must be taken into account. Additionally, the model only allows for individuals to become reinfected at the rate of β, which may not be the case for all reinfections. Some individuals may become reinfected multiple times, in which case their infection parameters would likely have different values.

The viral memetic model proposed here is a simplistic first step in modeling infectious outbreaks of culture and ideas.

4. Conclusion

A viral memetic model has been developed based on the modified SIR model of infectious diseases. The model has been analyzed and has been verified according to historic Internet search data. The model provides excellent fits for data sets exhibiting a sharp initial spike followed by a gradual taper.

It is interesting to see that a process as apparently random as the self-propagation of an idea through the minds of the masses can be modelled by such a simple model and fit a theoretical curve so closely. Since the SIR model was a basis for this work, the success of the modelling indicates that memes can indeed be considered to be infectious in nature, which opens up whole new frontiers for the likes of education, marketing, and even politics.

Acknowledgments

The first author was partially supported by NSERC of Canada and MITACS. The authors thank the referee for very valuable comments.

References

[1] R. Dawkins, The Selfish Gene, Oxford University Press, 1976.
[2] J. Leskovec, L. Beckstrom, and J. Kleinberg, Memetracking and the dynamics of the news cycle, Technical report, Cornell University and Stanford University, 2009.
[3] F. Débarre, SIR models epidemics, Technical report, ETHZurich, 2008.
[4] M. Kubo, K. Naruse, H. Satio and T. Matubara, The possibility of an epidemic meme analogy for web community population analysis, Technical report, National Defense Academy of Japan, 2007.
[5] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. R. Soc. Lond. B Biol. Sci. 115 (1927) 700–721.
[6] L.M.A. Bettencourt, A. Cintrón-Aias, D.I. Kaiser, C. Castillo-Chávez, The power of a good idea: quantitative modeling of the spread of ideas from epidemiological models, Physica A 364 (2006) 513–536.
[7] K. Kawachi, Deterministic models for rumor transmission, Nonlinear Anal.: Real World Appl. 9 (2008) 1989–2028.
[8] F.J. Santonjaa, A.C. Tarazonaa, R.J. Villanueva, A mathematical model of the pressure of an extreme ideology on a society, Comput. Math. Appl. 56 (2008) 836–846.
[9] S.H. Strogatz, Nonlinear Dynamics and Chaos, Westview Press, 2001.
[10] http://www.google.com/intl/en/trends/about.html.