Collective Attention on the Web

Christian Bauckhage
University of Bonn
Fraunhofer IAIS
christian.bauckhage@iais.fraunhofer.de

Kristian Kersting
TU Dortmund University
kristian.kersting@cs.tu-dortmund.de

the essence of knowledge
Boston — Delft

Foundations and Trends® in Web Science

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

C. Bauckhage and K. Kersting. *Collective Attention on the Web*. Foundations and Trends® in Web Science, vol. 5, no. 1-2, pp. 1–136, 2014.

This Foundations and Trends® issue was typeset in LaTeX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-68083-205-1

Foundations and Trends® in Web Science

Volume 5, Issue 1-2, 2014

Editorial Scope

Topics

Foundations and Trends® in Web Science publishes survey and tutorial articles in the following topics:

- Agents and the semantic web
- Collective intelligence
- Content management
- Databases on the web
- Data mining
- Democracy and the web
- Dependability
- Economics of information and the web
- E-crime
- E-government
- Emergent behaviour
- Ethics
- Hypertext/Hypermedia
- Identity
- Languages on the web
- Memories for life
- Mobile/Pervasive
- Network infrastructures
- Performance
- Privacy
- Scalability
- Security
- Semantic web
- Social networking
- Standards
- The law and the web
- The web as an educational tool
- The web in the developing world
- Trust and provenance
- Universal usability
- User interfaces
- Virtual reality
- Web art
- Web governance
- Search
- Web services

Information for Librarians

Foundations and Trends® in Web Science, 2014, Volume 5, 4 issues. ISSN paper version 1555-077X. ISSN online version 1555-0788. Also available as a combined paper and online subscription.

Foundations and Trends® in Web Science
Vol. 5, No. 1-2 (2014) 1–136

DOI: 10.1561/1800000024

Collective Attention on the Web

Christian Bauckhage
University of Bonn
Fraunhofer IAIS
christian.bauckhage@iais.fraunhofer.de

Kristian Kersting
TU Dortmund University
kristian.kersting@cs.tu-dortmund.de

Contents

Abstract

Understanding the dynamics of collective human attention has been called a key scientific challenge for the information age. Tackling this challenge, this monograph explores the dynamics of collective attention related to Internet phenomena such as Internet memes, viral videos, or social media platforms and Web-based businesses. To this end, we analyze time series data that directly or indirectly represent how the interest of large populations of Web users in content or services develops over time. Regardless of regional or cultural contexts, we generally observe strong regularities in time series that reflect attention dynamics and we discuss mathematical models that provide plausible explanations as to what drives the apparently dominant dynamics of rapid initial growth and prolonged decline.

C. Bauckhage and K. Kersting. *Collective Attention on the Web*. Foundations and Trends® in Web Science, vol. 5, no. 1-2, pp. 1–136, 2014.
DOI: 10.1561/1800000024.

1

Introduction

The Web has evolved to be many things: a network, a service, a means of social interaction, a marketplace, a source of news, a repository of knowledge, a database of multimedia content, and an integral part of human activity. This raises the need to investigate how the Web will further evolve as a social-, commercial-, and technical platform and has given rise to the new discipline of Web science [Berners-Lee et al., 2006, Hendler et al., 2008].

By its very nature, Web science is an interdisciplinary endeavor that involves aspects of sociology, psychology, economics, computer science, and data science and the content of this monograph illustrates this. We will investigate an apparently widespread sociological or psychological Web phenomenon from the point of view of data science. In particular, we will discuss how to harness data scientific methods in order to develop an understanding of the dynamics of collective attention on the Web. In fact, our motivation behind this monograph is perfectly summarized by the following quote:

> The subject of collective attention is central to an information age where millions of people are inundated with daily messages. . . . It is thus of interest to understand how atten-

> tion to novel items propagates and eventually fades among large populations. [Wu and Huberman, 2007]

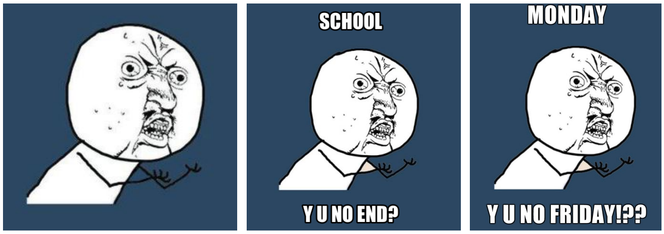
We first encountered the peculiar dynamics of collective attention processes when we began studying Internet memes a couple of years ago [Bauckhage, 2011].

The term *Internet meme* refers to the phenomenon of content that spreads rapidly among Web users. It alludes to a theory by Dawkins [1976] who postulated *memes* as a cultural analogon of biological genes so as to explain how rumors, catch-phrases, melodies, or fashion trends replicate through a population.

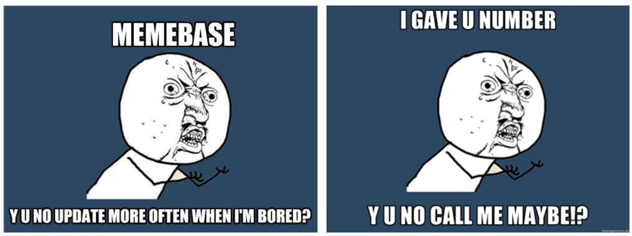
Correspondingly, Internet memes are catch phrases or humorous or repugnant pictures or video clips that "go viral" on the Web. While the phenomenon of viral content can be traced back to the early days of the Web, it is because of the interactive and participatory nature of modern social media such as content sharing platforms or social networking sites that Internet memes have become a staple of contemporary Web culture. They typically originate from platforms like *4chan*, *tumblr*, or *youtube*, gain notoriety via social news and entertainment sites such as *reddit*, *failblog*, *memegenerator*, or *quickmeme* and then spread through the social Web at large [Bauckhage, 2011, Coscia, 2013, Shifman, 2013].

Internet memes are *dynamic media objects* that evolve through commentary or parody. Consider, for example, the "y u no" meme shown in Figure 1.1. It first appeared on *tumblr* in 2010 and quickly found its way to *memegenerator* from where it spread virally. In its basic form, the meme consists of an image of a stick figure whose angry face was copied from the Japanese anime series Gantz (Figure 1.1(a)). It typically contains a text in short messaging style that poses mundane questions as to modern life and culture (Figure 1.1(b)). Mutations include self-referential variants that allude to *meme culture* (Figure 1.1(c)) as well as versions that deviate from the original phenotype (Figure 1.1(d)). Also, the meme occasionally occurs in media outside of the Internet but is then reported back on the Web, for instance on social networking sites (Figure 1.1(e)). Internet memes therefore transgress media and cultural boundaries and can be characterized as inside jokes that many people are in on.

Full text available at: http://dx.doi.org/10.1561/1800000024



(a) the “y u no” guy (b) instances of the “y u no” meme

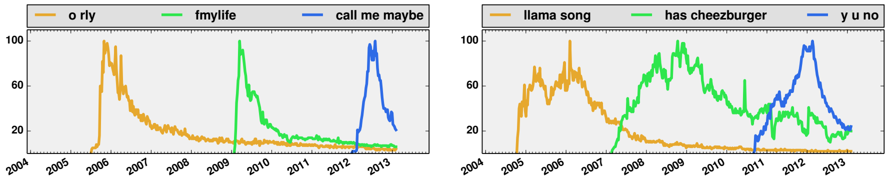


(c) the “y u no” meme with references to the Web site *memebase* and to the “call me maybe” meme

(d) mutations of the “y u no” meme alluding to pop cultural items such as an anime movie or a game franchise

(e) the “y u no” meme appearing on a birthday cake and on the front page of a printed news paper

Figure 1.1: Example of an Internet meme. Instances of the “y u no” meme consist of a simple image macro and a grammatically carefree piece of text that calls to attention questions of everyday life and contemporary culture. The meme first appeared on *tumblr* in 2010; as of this writing, querying the Google search engine for “y u no” yields more than a million results.



(a) positively skewed, narrow peaks, rather long tails

(b) positively skewed, broad peaks, rather short tails

Figure 1.2: Prototypical examples of meme related time series retrieved from Google Trends. While details are chaotic, there appear to be common, general trends as to how collective interest in individual Internet memes grows and declines.

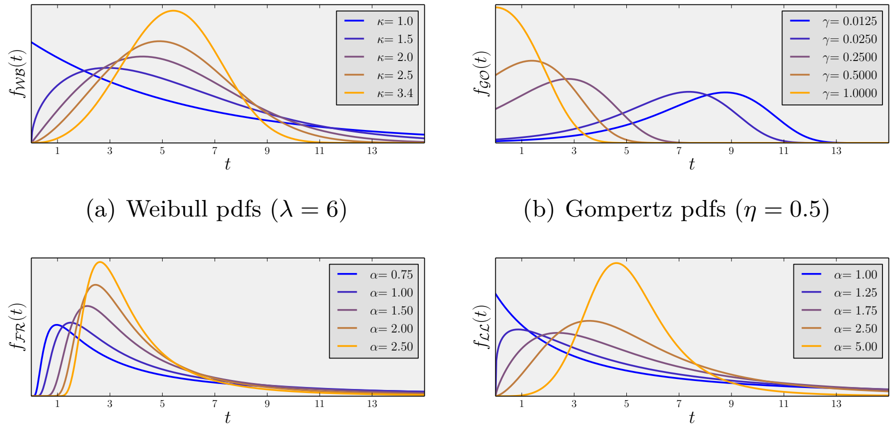
In addition to their content dynamics, Internet memes also show characteristic properties regarding their life cycles. While some were observed to go in and out of popularity in just a matter of weeks, others attract collective attention for extended periods of time. This is exemplified in Figure 1.2 which shows meme related time series retrieved from Google Trends. The graphs indicate how worldwide interest in individual memes (measured in terms of relative search frequencies) grew and declined over time. Although the short term dynamics of these time series appear chaotic, there are characteristic *general trends*: after a point of onset, public interest in a meme grows rapidly but once a meme has reached peak popularity, interest begins to fade more or less quickly.

Interestingly, attention dynamics like these are not exclusive to the phenomenon of Internet memes. Among others, we also observed them to manifest in

- Web search frequency data related to buzz words from the area of information technology [Bauckhage et al., 2013a]

- Web search frequency data related to social media services or e-commerce websites [Bauckhage et al., 2014]

- time series of view counts of popular *youtube* videos [Bauckhage et al., 2015a]



(a) Weibull pdfs ($\lambda = 6$)

(b) Gompertz pdfs ($\eta = 0.5$)

(c) Frechet pdfs ($\beta = 3$)

(d) Log-Logistic pdfs ($\beta = 5$)

Figure 1.3: Examples of probability density functions (pdfs) whose shapes resemble those of the empirical time series in Figure 1.2.

- time series indicating daily playing times people spent on online games [Bauckhage et al., 2012]
- time series reflecting the buying behaviors of players of freemium games [Sifa et al., 2015].

Taken together, all these observations suggest that the dynamics of collective attention on the Web seem to be governed by common latent principles or processes. The obvious question therefore is, if data collected from the Web will allow us to identify or at least to reliably characterize the nature of these hidden, *i.e.* not directly observable, processes.

Seen from the point of view of data science, answering this question seems but a mere exercise in model fitting. Indeed, there are well established scientific tools for time series analysis and it should be easy to fit more or less flexible mathematical models such as shown in Figure 1.3.

However, the crucial point we are trying to bring forward in this monograph is that *the problem we are dealing with is first and foremost a problem of model selection rather than a problem of mere model fitting.*

To clarify this claim, we note that, when properly parametrized, the examples of various probability distributions shown in Figure 1.3 all seem to be able to account for the general behavior of the time series in Figure 1.2. In other words, it will generally be no problem whatsoever to devise more or less sophisticated mathematical functions that capture the attention dynamics we are interested in. The real problem is whether every such model makes sense, or, to paraphrase once more, whether every such model is plausibly interpretable in terms of *physical* processes.

Note that by using the term *physical*, we do not necessarily refer to mechanisms studied in *physics* but express the fact that a convincing mathematical model of Web phenomena should not just fit observed data but should also allow for explaining them in terms of concepts that are grounded in the real world. In our context, this is to say that any model of the dynamics of collective attention on the Web should only be deemed appropriate if it can be tied in with psychological or social phenomena.

Alas, much of the literature on data analytics for Web science is agnostic of this reasonable requirement; rather, the focus often seems to be on the aspect of goodness of fit than on the aspect of plausibility. This has previously been noted by Lazer et al. [2014] who vehemently criticized the lack of interpretability and the "big data hubris" of purely data driven approaches to Web data analytics for their potential of over-fitting and misleading results.

Given these preliminaries, we can summarize our contributions in this monograph as follows: We are interested in the temporal dynamics of collective attention on the Web and our object of study are discrete time series such as in Figure 1.2 which reflect how the interest of large populations of Web users in a topic evolves over longer periods of time. We analyze data like these using mainly statistical tools, however, our methodology follows a model driven rather than a data driven paradigm and therefore adheres to the criticism brought forth by Lazer et al. [2014].

To be more specific, we explore to what extent collective attention dynamics on the Web can be understood in terms of

- growth processes that are known from the study of fads in fields like sociology or cultural studies
- diffusion processes that are studied by economists trying to understand the diffusion of innovations or goods
- spreading processes that are known to occur in (social) networks and are frequently studied in physics
- epidemic processes that characterize viral outbreaks and are of major interest in medicine and mathematical epidemiology.

This is to say that we will propose and investigate different models that can account for the noticeably skewed nature of attention related time series such as in Figure 1.2. Admittedly, this will be a mathematical endeavor but in order not to lose our focus on the topic of collective attention on the Web, we defer more technical material to the appendix. Nevertheless, we very much encourage our mathematically inclined readers to work through the appendix so as to fully appreciate the depth of the approaches we consider.

References

D. Acemoglu, A. Ozdaglar, and M. Yildiz. Diffusion of Innovations in Social Networks. In *Proc. Int. Conf. on Decision and Control*. IEEE, 2011.

A. Acerbi, S. Ghirlanda, and M. Enquist. The Logic of Fashion Cycles. *PLoS ONE*, 7(3):e32541, 2012.

E. Adar and A. Adamic. Tracking Information Epidemics in Blogspace. In *Proc. Int. Conf. on Web Intelligence*. IEEE/WIC/ACM, 2005.

C. Artola and E. Galan. Tracking the Future on the Web: Construction of Leading Indicators using Internet Searches. Documentos Ocasionales 1203, Banco de Espana, 2012.

S. Asur, B. Huberman, G. Zsabo, and C. Wang. Trends in Social Media: Persictence and Decay. In *Proc. ICWSM*. AAAI, 2011.

A. Barabasi and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, 1999.

M. Barthelemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani. Velocity and Hierarchical Spread of Epidemic Outbreaks in Scale-free Networks. *Physical Review Letters*, 92(17):178701, 2004.

F. Bass. A New Product Growth Model for Consumer Durables. *Management Science*, 15(5):215–227, 1969.

C. Bauckhage. Insights into Internet Memes. In *Proc. ICWSM*. AAAI, 2011.

C. Bauckhage. Computing the Kullback-Leibler Divergence between two Weibull Distributions. *arXiv:1310.3713 [cs.IT]*, 2013.

C. Bauckhage. Characterizations and Kullback-Leibler Divergence of Gompertz Distributions. *arXiv:1402.3193 [cs.IT]*, 2014a.

C. Bauckhage. Computing the Kullback-Leibler Divergence between two Generalized Gamma Distributions. *arXiv:1401.6853 [cs.IT]*, 2014b.

C. Bauckhage, K. Kersting, R. Sifa, C. Thurau, A. Drachen, and A. Canossa. How Players Lose Interest in Playing a Game: An Empirical Study Based on Distributions of Total Playing Times. In *Proc. Conf. on Computational Intelligence in Games*. IEEE, 2012.

C. Bauckhage, K. Kersting, and F. Hadiji. Mathematical Models of Fads Explain the Temporal Dynamics of Internet Memes. In *Proc. ICWSM*. AAAI, 2013a.

C. Bauckhage, K. Kersting, and B. Rastegarpanah. The Weibull as a Model of Shortest Path Distributions in Random Networks. In *Proc. Int. Workshop on Mining and Learning with Graphs*, Chicago, IL, USA, 2013b. ACM.

C. Bauckhage, K. Kersting, and B. Rastegarpanah. Collective Attention to Social Media Evolves According to Diffusion Models. In *Proc. WWW*. ACM, 2014.

C. Bauckhage, F. Hadiji, and K. Kersting. How Viral Are Viral Videos? In *Proc. ICWSM*. AAAI, 2015a.

C. Bauckhage, K. Kersting, and F. Hadiji. Parameterizing the Distance Distribution of Undirected Networks. In *Proc. UAI*, 2015b.

A. Bemmaor. Modeling the Diffusion of New Durable Goods: Word-of-mouth Effect Versus Consumer Heterogeneity. In G. Laurent, G.L. Lilien, and B. Pras, editors, *Research Traditions in Marketing*, pages 201–229. Springer, 1994.

T. Berners-Lee, W. Hall, J. Hendler, K. O'Hara, N. Shadbolt, and D. Weitzner. A Framework for Web Science. *Foundations and Trends in Web Science*, 1(1):1–130, 2006.

M. Bernstein, A. Monroy-Hernandez, D. Harry, P. Andr, K. Panovich, and G. Vargas. 4chan and /b/: An Analysis of Anonymity and Ephemerality in a Large Online Community. In *Proc. ICWSM*. AAAI, 2011.

S. Bikhchandani, D. Hirshleifer, and I. Welch. Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades. *J. of Economic Perspectives*, 12(3):151–170, 1998.

D. Bild, Y. Liu, R. Dick, Z. Morley Mao, and D. Wallach. Aggregate Characterization of User Behavior in Twitter and Analysis of the Retweet Graph. *arXiv:1402.2671[cs.SI]*, 2014.

V. Blondel, J. Guillaume, J. Hendrickx, and R. Jungers. Distance Distribution in Random Graphs and Application to Network Exploration. *Physical Review E*, 76(6):066101, 2007.

I. Bordino, S. Battiston, G. Caldarelle, M. Cristelli, A. Ukkonen, and I. Weber. Web Search Queries can Predict Stock Market Volumes. *PLoS ONE*, 7(7): e40014, 2012.

T. Britton. Stochastic Epidemic Models: A Survey. *Mathematical Biosciences*, 225(1):24–35, 2010.

T. Broxton, Y. Interian, J. Vaver, and M. Wattenhofer. Catching a Viral Video. *J. of Intelligent Information Systems*, 40(2):241–259, 2013.

C. Budak, D. Agrawal, and A. El Abbadi. Limiting the Spread of Misinformation in Social Networks. In *Proc. WWW*. ACM, 2010.

J. Burgess. All Your Chocolate Rain Are Belong To Us? Viral Video, YouTube and the Dynamics of Participatory Culture. In G. Lovink and S. Niederer, editors, *The VideoVortex Reader*. Institute of Network Cultures, Amsterdam, 2008.

J. Cannarella and J. Spechler. Epidemiological Modeling of Online Social Network Dynamics. *arXiv : 1401.4208 [cs.SI]*, 2014.

J. Castle, N. Fawcett, and D. Hendry. Nowcasting Is Not Just Comtemporaeneous Forecasting. *National Institute Economic Review*, 210(1):71–89, 2009.

H. Choi and H. Varian. Predicting the Present with Google Trends. *Economic Record*, 88(S1):2–9, 2012.

L. Christiansen, T. Schimoler, R. Burke, and B. Mobasher. Modeling Topic Trends on the Social Web Using Temporal Signatures. In *Proc. WIDM*. ACM, 2012.

A. Cintron-Arias. To Go Viral. *arXiv:1402.3499 [physics.soc-ph]*, 2014.

R. Cohen and S. Havlin. *Complex Networks*. Cambridge University Press, 2010.

M. Coscia. Competition and Success in the Meme Pool: A Case Study on Quickmeme.com. In *Proc. ICWSM*. AAAI, 2013.

M. Coscia. Average is Boring: How Similarity Kills a Meme's Success. *Scientific Reports*, 4(6477):1–7, 2014.

R. Crane and D. Sornette. Robust Dynamic Classes Revealed by Measuring the Response Function of a Social System. *Proc. of the National Academy of Sciences*, 105(41):15649–15653, 2008a.

R. Crane and D. Sornette. Viral, Quality, and Junk Videos on YouTube: Separating Content from Noise in an Information-Rich Environment. In *Proc. Spring Symp. on Social Information Processing*. AAAI, 2008b.

G. Crooks. The Amoroso Distribution. *arXiv:1005.3274 [math.ST]*, 2010.

Z. Da, J. Engelberg, and P. Gao. In Search of Attention. *J. of Finance*, 66 (5):1461–1499, 2011.

R. Dawkins. *The Selfish Gene*. Oxford University Press, 1976.

L. de Haan and A. Ferreira. *Extreme Value Theory*. Springer, 2006.

K. Dietz. Epidemics and Rumors: A Survey. *J. of the Royal Statistical Society A*, 130(4):505–528, 1967.

Y. Dover, J. Goldberg, and D. Shapira. Network Traces on Penetration: Uncovering Degree Distribution from Adoption Data . *Marketing Science*, 31(4):689–712, 2012.

F. Figueiredo, F. Benevenuto, and J. Almeida. The Tube over Time: Characterizing Popularity Growth of YouTube Videos. In *Proc. WSDM*. ACM, 2011.

A. Fronczak, P. Fronczak, and J. Holyst. Average Path Length in Random Networks. *Physical Review E*, 70(5):056110, 2004.

A. Gerow and M. Keane. Mining the Web for the Voice of the Herd to Track Stock Market Bubbles. In *Proc. IJCAI*. AAAI, 2011.

J. Ginsberg, M. Mohebbi, R. Patel, L. Brammer, M. Smolinski, and L. Brilliant. Detecting Influenza Epidemics Using Search Engine Query Data. *Nature*, 457(7232):1012–1014, 2009.

L. Gleser and D. Moore. The Effect of Dependence on Chi-Squared and Empiric Distribution Tests of Fit. *Annals of Statistics*, 11(4):1100–1108, 1983.

S. Goel, D. Watts, and D. Goldstein. The Structure of Online Diffusion Networks. In *Proc. EC*. ACM, 2012.

M. Gomez-Rodriguez, J. Leskovec, and B. Schölkopf. Structure and Dynamics of Information Pathways in Online Media. In *Proc. WSDM*. ACM, 2013.

L. Granka. Inferring the Public Agenda from Implicit Query Data. In *Proc. SIGIR*. ACM, 2009.

M. Granovetter. Threshold Models of Collective Behavior. *The American J. of Sociology*, 83(6):1420–1443, 1978.

L. Grossman. How to get famous in 3500 seconds. Time Magazine, April 24 2006.

J. Hendler, N. Shadbolt, W. Hall, T. Berners-Lee, and D. Weitzner. Web Science: An Interdisciplinary Approach to Understanding the Web. *Communications of the ACM*, 51(7):60–69, 2008.

J. Hermann, W. Rand, B. Schein, and N. Vedopivec. An Agent-Based Model of Urgent Diffusion in Social Media. Technical report, Social Science Research Network, 2013. http://dx.doi.org/10.2139/ssrn.2297167.

B. Huberman, P. Pirolli, J. Pitkow, and R. Lukose. Strong Regularities in World Wide Web Surfing. *Science*, 280(5360):95–97, 1998.

J. Iribarren and E. Moro. Impact of Human Activity Patterns on the Dynamics of Information Diffusion. *Physical Review Letters*, 103(3):038702, 2009.

P. Jain, J. Manweiler, and A. Acharya R. Choudhury. Scalable Social Analytics for Live Viral Event Prediction. In *Proc. ICWSM*. AAAI, 2014.

E. Jaynes. Information Theory and Statistical Mechanics. *Physical Review*, 106:620–630, 1957.

E. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

R. Jennrich and R. Moore. Maximum Likelihood Estimation by Means of Nonlinear Least Squares. In *Proc. of the Statistical Computing Section*. American Statistical Association, 1975.

L. Jiang, Y. Miao, Y. Zhang, Z. Lan, and A. Hauptmann. Viral Video Style: A Closer Look at Viral Videos on YouTube. In *Proc. Int. Conf. on Multimedia Retrieval*. ACM, 2014.

K. Joseph, J. Wintoki, and Z. Zhang. Forecasting Abnormal Stock Returns and Trading Volume Using Investor Sentiment: Evidence from Online Search. *Int. J. of Forecasting*, 27(4):1116–1127, 2011.

M. Keeling and K. Eames. Networks and Epidemic Models. *J. Royal Society Interface*, 2(4):295–307, 2005.

D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the Spread of Influence through A Social Network. In *Proc. KDD*. ACM, 2003.

C. Kleiber and S. Kotz. *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley, 2003.

M. Knobel and C. Lankshear. Online Memes, Affinities, and Cultural Production. In M. Knobel and C. Lankshear, editors, *A New Literacies Sampler*, chapter 9. Peter Lang Publishing, 2007.

A. Koch. The Logarithm in Biology: 1. Mechanisms Creating Log-Normal Distributions Exactly. *J. of Theoretical Biology*, 12(2):27–290, 1966.

B. Krausz and C. Bauckhage. Loveparade 2010: Automatic Video Analysis of a Crowd Disaster. *Computer Vision and Image Understanding*, 116(3): 307–319, 2012.

M. Kubo, K. Naruse, H. Sato, and T. Matubara. The Possibility of an Epidemic Meme Analogy for Web Community Population Analysis . In *Proc. Int. Conf. on Intelligent Data Engineering and Automated Learning*, 2007.

J. Kunegis. KONECT: the Koblenz Network Collection. In *Proc. WWW*. ACM, 2013.

J. Lawless. *Statistical Models and Methods for Lifetime Data*. Wiley, 2003.

D. Lazer, R. Kennedy, G. King, and A. Vesignani. The Parable of Google Flu: Traps in Big Data Analysis. *Science*, 343(6176):1203–1205, 2014.

J. Lehmann, B. Goncalves, J.J. Ramasco, and C. Cattuto. Dynamical Classes of Collective Attention in Twitter. In *Proc. WWW*. ACM, 2012.

J. Lescovec and E. Horvitz. Planetary-Scale Views on a Large Instant-Messaging Network. In *Proc. WWW*. ACM, 2008.

J. Leskovec, L. Adamic, and B. Huberman. The Dynamics of Viral Marketing. *ACM Transactions on the Web*, 1(1):5, 2007a.

J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, and M. Hurst. Patterns of Cascading Behavior in Large Blog Graphs. In *Proc. SDM*. SIAM, 2007b.

J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the Dynamics of the News Cycle. In *Proc. KDD*. ACM, 2009.

C. Lin, B. Zhao, Q. Mei, and J. Han. PET: A Statistical Model for Popular Events Tracking in Social Communities. In *Proc. KDD*. ACM, 2010.

C. Liu, R. White, and S. Dumais. Understanding Web Browsing Behavior through Weibull Analysis of Dwell Times. In *Proc. SIGIR*. ACM, 2010.

A. Lloyd and R. May. How Viruses Spread Among Computers and People. *Science*, 292(5520):1316–1317, 2001.

D. Luu, E. Lim, T. Hoang, and F. Chua. Modeling Diffusion in Social Networks Using Network Properties. In *Proc. ICWSM*. AAAI, 2012.

N. McLaren and R. Shanbhogue. Using Internet Search Data as Economic Indicators. *Bank of England Quarterly Bulletin*, 51(2):134–140, 2011.

J. Mellon. Search Indices and Issue Salience: the Properties of Google Trends as a Measure of Issue Salience. Sociology Working Papers 2011-01, University of Oxford, 2011.

R. Meyerson and E. Katz. Notes on a Natural History of Fads. *American J. of Sociolog*, 62(6):594–601, 1957.

M. Mitzenmacher. A Brief History of Generative Models for Power Law and Lognormal Distributions. *Internet Mathematics*, 1(2):226–251, 2004.

M. Newman. The Spread of Epidemic Diseases on Networks. *Physical Review E*, 66(1):016128, 2002.

E. Page. Continuous Inspection Scheme. *Biometrika*, 41(1–2):100–115, 1954.

R. Pastor-Satorras and A. Vespignani. Epidemic Spreading in Scale-Free Networks. *Physical Review Letters*, 86(14):3200–3203, 2001.

D. Pennock, G. Flake, S. Lawrence, E. Glover, and C. Gilles. Winners Don't Take All: Characterizing the Competition for Links on the Web. *Proc. of the National Academy of Sciences*, 99(8):5207–5211, 2002.

P. Phillips. Time Series Regression with a Unit Root. *Econometrica*, 55(2): 277–301, 1987.

J. Pinto, J. Almeida, and M. Goncalves. Using Early View Patterns to Predict the Popularity of YouTube Videos. In *Proc. WSDM*. ACM, 2013.

D. Pogue. Internet Memes 101: A Guide to Online Wackiness. New York Times, September 8, 2011.

B. Ribeiro. Modeling and Predicting the Growth and Death of Membership-Based Websites. In *Proc. WWW*. ACM, 2014.

H. Rinne. *The Weibull Distribution*. Chapman & Hall / CRC, 2008.

E. Robertson. Kickstarting Memes and Movements: A Distribution-Based Threshold Model. retrieved from: santafe.edu/media/, 2016.

D. Scharfstein and J. Stein. Herd Behavior and Investment. *American Economic Review*, 80(3):465–479, 1990.

D. Shamma, J. Yew, L. Kennedy, and E. Churchill. Viral Actions: Predicting Video View Counts Using Synchronous Sharing Behaviors. In *Proc. ICWSM*. AAAI, 2011.

X. Shao and Z. Zhang. Testing for Change Points in Time Series. *J. American Statistical Association*, 105(491):1228–1240, 2010.

L. Shifman. An Anatomy of a YouTube Meme. *New Media & Society March*, 14(2):187–203, 2012.

L. Shifman. *Memes in Digital Culture*. MIT Press, 2013.

R. Sifa, F. Hadiji, J. Runge, A. Drachen, K. Kersting, and C. Bauckhage. Predicting Purchase Decisions in Mobile Free-to-Play Games. In *Proc. AIIDE*. AAAI, 2015.

D. Southgate, N. Westoby, and G. Page. Creative Determinants of Viral Video Viewing. *Int. J. of Advertising*, 29(3):349–368, 2010.

E. Stacy. A Generalization of the Gamma Distribution. *The Annals of Mathematical Statistics*, 33(3):1187–1192, 1962.

T. Tassier. A Model of Fads, Fashions, and Group Formation. *Complexity*, 9 (5):51–61, 2003.

J. Teevan, D. Liebling, and G. Geetha. Understanding and Predicting Personal Navigation. In *Proc. WSDM*. ACM, 2011.

J. Thom and D. Millen. Stuff IBMers Say: Microblogs as an Expression of Organizational Culture. In *Proc. ICWSM*. AAAI, 2012.

A. Ukkonen. Indirect Estimation of Shortest Path Distributions with Small-World Experiments. In *Proc. Advances in Intelligent Data Analysis*, 2014.

A. Vazquez. Polynomial Growth in Branching Processes with Diverging Reproduction Number. *Physical Review Letters*, 96(3):038702, 2006.

D. Wang, C. Song, and A. Barabasi. Quantifying Long-Term Scientific Impact. *Scienc*, 342(6154):127–132, 2013.

L. Weng. *Information Diffusion on Online Social networks*. PhD thesis, Indiana University, 2014.

C. Winsor. The Gompertz Curve as a Growth Curve. *Proc. of the National Academy of Sciences*, 18(1):1–8, 1932.

F. Wu and B. Huberman. Novelty and Collective Attention. *Proc. of the National Academy of Sciences*, 104(45):17599–17601, 2007.

J. Yang and J. Leskovec. Patterns of Temporal Variation in Online Media. In *Proc. WSDM*. ACM, 2011.

B. Zeide. Analysis of Growth Equations. *Forrest Science*, 39(3):594–616, 1993.

L. Zhang, J. Zhao, and K. Xu. Who Creates Trends in Online Social Media: The Crowd or Opinion Leaders? *J. of Computer-Mediated Communiation*, 21(1):1–16, 2016.

C. Zorn. Modeling Duration Dependence. *Political Analysis*, 8(4):367–380, 2000.