

A MODEL FOR THE SPREADING OF FAKE NEWS

HOSAM MAHMOUD,* *The George Washington University*

Abstract

We introduce a model for the spreading of fake news in a community of size n . There are $j_n = \alpha n - g_n$ active gullible persons who are willing to believe and spread the fake news, the rest do not react to it. We address the question ‘How long does it take for $r = \rho n - h_n$ persons to become spreaders?’ (The perturbation functions g_n and h_n are $o(n)$, and $0 \leq \rho \leq \alpha \leq 1$.) The setup has a straightforward representation as a convolution of geometric random variables with quadratic probabilities. However, asymptotic distributions require delicate analysis that gives a somewhat surprising outcome. Normalized appropriately, the waiting time has three main phases: (a) away from the depletion of active gullible persons, when $0 < \rho < \alpha$, the normalized variable converges in distribution to a Gumbel random variable; (b) near depletion, when $0 < \rho = \alpha$, with $h_n - g_n \rightarrow \infty$, the normalized variable also converges in distribution to a Gumbel random variable, but the centering function gains weight with increasing perturbations; (c) at almost complete depletion, when $r = j - c$, for integer $c \geq 0$, the normalized variable converges in distribution to a convolution of *two* independent generalized Gumbel random variables. The influence of various perturbation functions endows the three main phases with an infinite number of phase transitions at the seam lines.

Keywords: fake news; rumors; Gumbel distribution; phase transition

2010 Mathematics Subject Classification: Primary 60C05
Secondary 60F05

1. Spreading of fake news

In a community, fake news and the like is started by someone and spreads out through interactions between individuals of that community. There are *active gullible* persons who are willing to believe the fake news and participate in spreading it, and there are persons who hear the fake news and do not react to it, either because they are savvy enough to see through the falsehood or are simply uninterested in spreading news, even when they believe it. One sees this phenomenon too often these days in social media.

We assume that the size of the entire community is n (large), and there are initially $n - j$ persons who will not participate at any point in spreading the news, $j - 1$ active gullible persons (potential spreaders), and one person who manufactures the news. The number of those who will not participate at any point in spreading the news remains unchanged throughout, whereas the number of spreaders increases, say to k , at which point there are $j - k$ remaining gullible persons who have not yet heard the news.

Interactions between members of the community take place over time. If a spreader comes in contact with an active gullible person who has not yet heard the news, the active gullible

Received 22 March 2019; revision received 25 November 2019.

* Postal address: Department of Statistics, The George Washington University, Washington, D.C. 20052, U.S.A. Email address: hosam@gwu.edu

person is converted into a spreader. Any other interaction has no effect and the picture remains unchanged.

A question of interest is ‘How long does it take until a certain proportion of the community have become spreaders?’ The question is obviously of interest to the media, policy makers, and society at large.

There is quite extensive literature on the spreading of rumors. A classic model is the Maki–Thomson [8] model, in which informed participants eventually become stiflers not interested in spreading the rumor. The Maki–Thomson model is also considered as a scenario in epidemiology, sociology, and biosciences [5, 9, 14]. Most studies have focused on the spreading of rumors (or diseases) in continuous time [1, 2, 12, 15]. However, a few sources reverted to discrete-time rumor models [6, 11, 13]. Our discrete-time model is essentially different, as it has no migration out of the spreaders. Also, in our scheme the stiflers play a less active role than in the Maki–Thomson model. Furthermore, we take a distributional view and ask a different kind of question, meant to address the multiple phase transitions that occur over time.

2. Notation

Time (discrete) refers to the number of interactions, that is, the number of times two persons from the community come in contact. This is to be contrasted with the notion of discrete time in [13], where the epochs are taken only at the points of change.

The average and variance of the waiting time come in terms of the n th harmonic number of order k , namely $H_n^{(k)} = \sum_{i=1}^n 1/i^k$. The following are well-known asymptotic representations for the harmonic numbers of the first two orders:

$$\begin{aligned} H_n^{(1)} &= \ln n + \gamma + O\left(\frac{1}{n}\right), \\ H_n^{(2)} &= \frac{\pi^2}{6} + O\left(\frac{1}{n}\right), \end{aligned} \tag{1}$$

where $\gamma = 0.5772\dots$ is Euler’s constant. It is customary to suppress the order k when it is 1.

There are generalized harmonic numbers with additional parameters. We need a generalized form of the harmonic number of order two, namely

$$\psi(1, x) = \sum_{i=0}^{\infty} \frac{1}{(i+x)^2},$$

which is also called the trigamma function or Hurwitz zeta function $\zeta(2, x)$; see [16, pp. 268–269].

We use $\text{Geo}(p)$ to indicate a geometric random variable with parameter p . The asymptotic distributions contain a Gumbel distribution. The Gumbel distribution with location parameter $\theta \in \mathbb{R}$ and scale parameter $\beta > 0$ has density

$$f(x) = \frac{1}{\beta} e^{-\frac{x-\theta}{\beta}} e^{-e^{-\frac{x-\theta}{\beta}}}, \quad -\infty < x < \infty,$$

and its moment-generating function is $e^{\theta t} \Gamma(1 - \beta t)$, for $t < 1/\beta$. A Gumbel random variable from such a distribution will be called $\text{Gum}(\theta, \beta)$. There is a generalization by Ojo [10]

of $\text{Gum}(0, \beta)$, for which the moment-generating function is $\Gamma(s - \beta t) / \Gamma(s)$. We call such a variable $\text{Ojo}(s, \beta)$.

The symbols $\xrightarrow{\mathcal{D}}$ and \xrightarrow{P} stand for convergence in distribution and in probability, respectively.

3. Scope and results

Let $W_r = W_{r,n,j}$ be the waiting time until $r \leq j$ persons have become spreaders. We assume that the parameter n is quite large, and

$$r_n \leq j_n, \tag{2}$$

$$j_n = \alpha n - g_n, \tag{3}$$

$$r_n = \rho n - h_n, \tag{4}$$

where $0 \leq \rho \leq \alpha \leq 1$, and the perturbation functions g_n and h_n are $o(n)$. In view of (2)–(4), all dependence is on n and we write W_{r_n} , and often simply just W_r , for $W_{r,n,j}$. Likewise, we often write j for j_n . We take the initial number of active gullible people to be positive (otherwise there is no problem to analyze).

For instance, j might be $\lfloor 6 \ln n + \pi \ln \ln n + 2e \rfloor$ (a case where $\alpha = 0$), might be $\lfloor \frac{1}{3}n + 2\sqrt{n} \rfloor$ (a case where $\alpha = \frac{1}{3}$), or might be $\lceil n - \frac{21}{50} \ln n \rceil$ (a case where $\alpha = 1$). Also, r_n in (4) has a similar interpretation. The case $\alpha = 0$ and $\rho = 0$ presents a less interesting degeneracy that is briefly discussed below in interleaved remarks.

Later on, we develop the exact moment-generating function of W_r , from which we can extract the exact mean and variance by taking derivatives. However, it is more straightforward to find the exact mean and variance, and consequently their asymptotic equivalents, from basic properties of independent geometric random variables.

One main result is the following.

Theorem 3.1. *In a community of size n , in which one person starts fake news, there are $j_n - 1$ active gullible persons and the remaining persons will never participate in spreading the fake news. Let W_{r_n} be the waiting time until r_n persons have become spreaders. Assume j_n and r_n follow (2)–(4). If $0 < \rho < \alpha$, as $n \rightarrow \infty$ we have*

$$\frac{W_{r_n} - \frac{1}{2\alpha(1-\frac{g_n}{\alpha n})} n \ln n}{n} \xrightarrow{\mathcal{D}} \text{Gum}\left(\frac{1}{2\alpha} \ln\left(\frac{\alpha\rho}{\alpha-\rho}\right), \frac{1}{2\alpha}\right).$$

Interpretation 3.1. The term $1/(1 - g_n/(\alpha n))$ can be expanded into an infinite series. If $g(n) = \Omega(n/(\ln n)^{1/(k-1)})$ but $g_n = o(n/(\ln n)^{1/k})$, the series can be stopped after k terms, and the tail can be claimed in an $o(n)$ remainder. But then, this remainder vanishes upon scaling by n . Slutsky’s theorem [7, p. 146] then assures us that we can truncate the series after k terms. For instance, j_n might be $\lceil \frac{2}{3}n - 2n^{\frac{5}{8}} + 3 \rceil$, i.e., we have $g_n = 2n^{\frac{5}{8}} + O(1)$, and $r_n \sim \frac{3}{7}n$, in which case the theorem takes the simple form

$$\frac{W_{r_n} - \frac{3}{4} n \ln n}{n} \xrightarrow{\mathcal{D}} \text{Gum}\left(\frac{3}{4} \ln\left(\frac{6}{5}\right), \frac{3}{4}\right).$$

Or, j_n might be $\lfloor \frac{3}{4}n + 5n \ln^{-\frac{3}{4}} n - n \ln^{-\frac{7}{9}} n + 2\sqrt[9]{n} + 19 \ln n - \ln \ln n + 13 \rfloor$, and $r_n \sim \frac{3}{5}n$, in which case we need two terms from the series. The theorem takes the form

$$\frac{W_{r_n} - \frac{2}{3} n \ln n - \frac{40}{9} n \ln^{\frac{1}{4}} n + \frac{8}{9} n \ln^{\frac{2}{9}} n}{n} = \frac{W_{r_n} - \frac{2}{3} (1 + \frac{4g_n}{3n}) n \ln n}{n} \xrightarrow{\mathcal{D}} \text{Gum}\left(\frac{2}{3} \ln 3, \frac{2}{3}\right).$$

More generally, If $g(n) = \Omega(n/(\ln n)^{1/(k-1)})$ but further we have $g_n = o(n/(\ln n)^{1/k})$, an equivalent form of Theorem 3.1 that captures the infinite number of fine phases is

$$\frac{W_{r_n} - \frac{1}{2\alpha} n \ln n - \frac{1}{2\alpha} n \ln n \sum_{i=1}^k (\frac{g_n}{\alpha n})^i}{n} \xrightarrow{\mathcal{D}} \text{Gum}\left(\frac{1}{2\alpha} \ln\left(\frac{\alpha\rho}{\alpha-\rho}\right), \frac{1}{2\alpha}\right).$$

Remark 3.1. The case $\rho = \alpha > 0$ presents a degeneracy in Theorem 3.1. An alternative statement to Theorem 3.1 is required.

Remark 3.2. The Gumbel distribution appears in the waiting time in coupon collection problems, too [3]. For a recent survey of coupon collection problems, see [4].

Theorem 3.2. *In a community of size n in which one person starts fake news, there are $j_n - 1$ active gullible persons and the remaining persons will never participate in spreading the fake news. Let W_{r_n} be the waiting time until r_n persons have become spreaders. Assume j_n and r_n follow (2)–(4). In the case of near depletion, with $\rho = \alpha > 0$ and $h_n - g_n \rightarrow \infty$, we have*

$$\frac{W_{r_n} - \frac{1}{\alpha(1-\frac{g_n}{\alpha n})} n \ln n - \frac{1}{\alpha(1-\frac{g_n}{\alpha n})} n \ln(h_n - g_n)}{n} \xrightarrow{\mathcal{D}} \text{Gum}\left(\frac{\ln \alpha}{\alpha}, \frac{1}{2\alpha}\right).$$

At almost complete depletion, where $r = j - c$ for integer $c \geq 0$, we have

$$\frac{W_{r_n} - \frac{1}{\alpha(1-\frac{g_n}{\alpha n})} n \ln n}{n} \xrightarrow{\mathcal{D}} \text{Ojo}\left(c + 1, \frac{1}{2\alpha}\right) + \text{Gum}\left(\frac{\ln \alpha}{\alpha}, \frac{1}{2\alpha}\right).$$

The Ojo and Gumbel random variables are independent.

The rest of this manuscript is organized as follows. In Section 4 we set up the waiting time as a convolution of geometric random variables. The convolution is used in a straightforward way in Section 5 to derive the mean and variance of the waiting time, both exactly and asymptotically. After the initial setup of the moment-generating function in Section 6, the asymptotics for the cases far from depletion, near depletion, and for almost complete depletion are separated in Subsections 6.1 and 6.2. The arguments in these subsections constitute proofs for Theorems 3.1 and 3.2. For æsthetic reasons, the case of complete depletion is treated separately in Subsection 6.3.

4. Waiting time

A Markov chain underlies the model. The status of the population can be specified by k , the number of spreaders. When the number of spreaders is k , there are $j - k$ gullible persons who have not yet heard the fake news, and as stated the number of persons uninterested in participating remains the same throughout. So, it is sufficient to label the state representing the condition of the population at this point with k . The initial state is 1.

The Markov chain advances from state k to state $k + 1$ with probability $p_{n,j,k} = k(j - k) / \binom{n}{2}$, and stays at state k with the probability $1 - p_{n,j,k}$. Thus, it takes a $\text{Geo}(p_{n,j,k})$ amount of time to advance from state k to state $k + 1$. The geometric random variables associated with the various states are independent. The waiting time until r persons are spreaders has a representation as a convolution of geometric random variables:

$$W_r = \sum_{k=1}^{r-1} \text{Geo}\left(\frac{k(j-k)}{\binom{n}{2}}\right).$$

5. Average and variance of waiting time

From the geometric convolution set up in Section 4, we find the following.

Proposition 5.1. *We have the average*

$$\mathbb{E}[W_r] = \frac{n(n-1)}{2j} (H_{r-1} + H_{j-1} - H_{j-r}).$$

As $n \rightarrow \infty$, we have

$$\mathbb{E}[W_r] \sim \begin{cases} \frac{H_{r-1}}{2\alpha} n, & \text{for fixed } r, \\ \frac{1}{2\alpha} n \ln r_n, & \text{for } r_n \rightarrow \infty \text{ and } r_n = o(n), \\ \frac{1}{2\alpha} n \ln n, & 0 < \rho < \alpha, \\ \frac{1}{\alpha} n (\ln n - \frac{1}{2} \ln (h_n - g_n)), & \text{for } r_n = \alpha n - h_n, h_n - g_n \rightarrow \infty, \\ \frac{1}{\alpha} n \ln n, & \text{for } r_n = j_n - c, c \geq 0. \end{cases}$$

Proof. The expected value of $\text{Geo}(p)$ is $1/p$, from which we find

$$\begin{aligned} \mathbb{E}[W_r] &= \binom{n}{2} \sum_{k=1}^{r-1} \frac{1}{k(j-k)} \\ &= \frac{n(n-1)}{2j} \sum_{k=1}^{r-1} \left(\frac{1}{k} + \frac{1}{j-k}\right) \\ &= \frac{n(n-1)}{2j} (H_{r-1} + H_{j-1} - H_{j-r}). \end{aligned}$$

Using the asymptotic equivalent in (1), we get the stated asymptotics. □

Remark 5.1. The asymptotic equivalent of the average at $r_n = j_n - c$ is twice that at $r_n \sim \rho n$, for $\rho \in (0, \alpha)$. The mathematical reason for this phase transition is that, for $0 < \rho < \alpha$, each of the three harmonic numbers provides an asymptotic value of $\ln n$. That is, $\ln n$ appears twice positively and once negatively. When $r_n = j_n - c$, the negative harmonic number is attenuated to $O(1)$. Thus, $\ln n$ appears only twice positively. Note how the function $\ln(h_n - g_n)$ fills the gap at the seam line between the cases $\rho < \alpha$ and the almost near completion, where $r_n = j_n - c$. This explains how the doubling of the average waiting time is not abrupt.

Remark 5.2. Of less interest is the case where $j_n = o(n)$. In this case, $E[W_{r_n}] = O(j_n^{-1}n^2 \ln j_n)$.

As in the mean of W_r , the variance of the waiting time exhibits phase transitions. The variance of $\text{Geo}(p)$ is $(1 - p)/p^2$. Whence, by the independence of the geometric random variables, we have

$$\text{Var}[W_r] = \sum_{k=1}^{r-1} \frac{1 - k(j - k)/\binom{n}{2}}{k^2(j - k)^2/\binom{n}{2}^2}.$$

After a somewhat tedious expansion of the summand into partial fractions, and summing each individual fraction as a harmonic number, we get the following.

Proposition 5.2. *We have the variance*

$$\begin{aligned} \text{Var}[W_r] = & \frac{\binom{n}{2}^2}{j^3} (H_{r-1} + 2H_{j-1} - 2H_{j-r}) \\ & + \frac{\binom{n}{2}^2}{j^2} (H_{r-1}^{(2)} + 2H_{j-1}^{(2)} - 2H_{j-r}^{(2)}) \\ & - \frac{\binom{n}{2}}{j} (H_{r-1} + 2H_{j-1} - 2H_{j-r}). \end{aligned}$$

Upon scaling, we have the limits

$$\frac{1}{n^2} \text{Var}[W_r] \sim \begin{cases} \frac{\pi^2 - 6\psi(1, r)}{24\alpha^2}, & r_n = r \text{ is constant,} \\ \frac{\pi^2}{24\alpha^2}, & r_n \rightarrow \infty \text{ and } r_n = o(n), \\ & \text{or } 0 < \rho < \alpha, \\ & \text{or } \rho = \alpha \text{ and } h_n - g_n \rightarrow \infty, \\ \frac{1}{24\alpha^2} (\pi^2 + 12\psi(1, h_n - g_n + 1)), & \rho = \alpha \text{ and } h_n - g_n = O(1), \\ \frac{\pi^2}{8\alpha^2}, & r_n = j_n \text{ (i.e., } \rho = \alpha \text{ and } h_n = g_n). \end{cases}$$

where $\psi(1, r + 1)$ is the series given in Section 2.

Proof. The steps leading to the exact variance were discussed before the statement of the theorem. Toward asymptotics, observe that all the terms involving first-order harmonics are only $O(n \ln n)$, whereas the second-order harmonics have the expansion

$$H_i = \frac{\pi^2}{6} - \psi(1, i + 1).$$

Furthermore, the function $\psi(1, i + 1)$ vanishes as $i \rightarrow \infty$. □

Remark 5.3. As in the mean, and for a similar reason, the asymptotic equivalent of the variance at $0 < \rho < \alpha$ is twice that near depletion, where $\rho = \alpha$.

For fixed r , the function $\psi(1, r + 1) = \sum_{s=1}^{\infty} 1/(s + r + 1)^2$ lowers the variance below that in the middle range $0 < \rho < \alpha$. When r approaches infinity, the ψ function begins to vanish, and the case merges into the middle range.

Remark 5.4. Barring the case of fixed r , the standard deviation in each of the other phases is of smaller order than the mean of the same phase. Therefore, we have a concentration law following from the Chebyshev inequality. Namely, we have $W_{r_n}/E[W_{r_n}] \rightarrow 1$ as $n \rightarrow \infty$.

6. Limit distributions

The geometric random variable $\text{Geo}(p)$ has moment-generating function $pe^t/(1 - (1 - p)e^t)$, so the convolution in W_r has the moment-generating function

$$\begin{aligned} \phi_r(t) &= \prod_{k=1}^{r-1} \frac{k(j-k)e^t}{\binom{n}{2} - \left(\binom{n}{2} - k(j-k)\right)e^t} \\ &= (-1)^{r-1} \prod_{k=1}^{r-1} \frac{k(j-k)}{k^2 - jk + \binom{n}{2}(1 - e^{-t})}. \end{aligned} \tag{5}$$

The asymptotic analysis of this moment-generating function is different for the near-depletion case $r = j - x_n$, with $x_n \rightarrow \infty$ while $x_n = o(n)$, from the asymptotics of the cases that are far away from depletion. Yet again, the case $r = j - c$ of near-complete depletion, that is $x_n = O(1)$, presents another phase change from the case of near depletion.

An effective way to deal with the representation (5) is to think of the denominator of each term of the product as a factored quadratic function of k . From this vantage point, we see that

$$\phi_r(t) = (-1)^{r-1} \prod_{k=1}^{r-1} \frac{k(j-k)}{(k - k_1(n, j, t))(k - k_2(n, j, t))},$$

where $k_1(n, j, t)$ and $k_2(n, j, t)$ are the roots of the equation

$$k^2 - jk + \binom{n}{2}(1 - e^{-t}) = 0,$$

that is, we have the roots

$$k_i(n, j, t) = \frac{1}{2}j \pm \frac{1}{2}\sqrt{j^2 - 2n(n-1)(1 - e^{-t})}, \quad \text{for } i = 1, 2.$$

(We take k_1 to be the root with the positive sign.) We can express the moment-generating function in terms of gamma functions:

$$\phi_r(t) = \frac{\Gamma(r)\Gamma(j)}{\Gamma(j-r+1)} \prod_{k=1}^{r-1} \frac{1}{(k_1(n, j, t) - k)(k - k_2(n, j, t))}.$$

We first develop an asymptotic equivalent for $E[e^{W_r u/n}]$. When t is small enough (as intended in the use of $t = u/n$, for fixed u and large n), we can work with accurate approximations of the roots. We have

$$\begin{aligned} k_1\left(n, j, \frac{u}{n}\right) &= \frac{1}{2}j + \frac{1}{2}j\sqrt{1 - \frac{2n(n-1)}{j^2}(1 - e^{-u/n})} \\ &= j - \frac{(n-1)}{2j}u + O\left(\frac{1}{n}\right). \end{aligned}$$

We are considering $j = \alpha n - g_n$, with $\alpha > 0$ and $g_n = o(n)$, in which case we get

$$\begin{aligned} k_1\left(n, j, \frac{u}{n}\right) &= j - \frac{n}{2j} u + O\left(\frac{1}{n}\right) \\ &= j - \frac{1}{2\alpha\left(1 - \frac{g_n}{\alpha n}\right)} u + O\left(\frac{1}{n}\right) \\ &= j - \frac{1 + b_n}{2\alpha} u + O\left(\frac{1}{n}\right), \end{aligned} \tag{6}$$

where $b_n = \sum_{i=1}^{\infty} (g_n/(\alpha n))^i$. Note that $b_n = O(g_n/n) \rightarrow 0$ as $n \rightarrow \infty$. Similarly, we have

$$k_2\left(n, j, \frac{u}{n}\right) = \frac{1 + b_n}{2\alpha} u + O\left(\frac{1}{n}\right). \tag{7}$$

6.1. Far away from depletion

Consider the case where $r_n \sim \rho n$, with $0 < \rho < \alpha$. So, there is still a linear number of active gullible persons and their conversion to spreaders will continue for a while. Thus, the case is far away from depleting the pool of active gullible persons.

In this case, we use (6) and (7) to write the asymptotic equivalent:

$$\begin{aligned} \phi_r\left(\frac{u}{n}\right) &= \frac{\Gamma(r) \Gamma(j)}{\Gamma(j - r + 1)} \prod_{k=1}^{r-1} \frac{1}{\left(j - \frac{1 + b_n}{2\alpha} u + O\left(\frac{1}{n}\right) - k\right)} \\ &\quad \times \prod_{k=1}^{r-1} \frac{1}{\left(k - \frac{1 + b_n}{2\alpha} u + O\left(\frac{1}{n}\right)\right)} \\ &= \frac{\Gamma(r) \Gamma(j)}{\Gamma(j - r + 1)} \times \frac{\Gamma\left(j - r + 1 - \frac{1 + b_n}{2\alpha} u + O\left(\frac{1}{n}\right)\right)}{\Gamma\left(j - \frac{1 + b_n}{2\alpha} u + O\left(\frac{1}{n}\right)\right)} \\ &\quad \times \frac{\Gamma\left(1 - \frac{1 + b_n}{2\alpha} u + O\left(\frac{1}{n}\right)\right)}{\Gamma\left(r - \frac{1 + b_n}{2\alpha} u + O\left(\frac{1}{n}\right)\right)}. \end{aligned}$$

The gamma function is continuous on \mathbb{R}^+ . As $b_n = O(g_n/n) \rightarrow 0$, it follows that

$$\Gamma\left(1 - \frac{1 + b_n}{2\alpha} u + O\left(\frac{1}{n}\right)\right) \rightarrow \Gamma\left(1 - \frac{1}{2\alpha} u\right) \quad \text{as } n \rightarrow \infty.$$

We take the remaining gamma functions in pairs, and to each pair we apply the Stirling approximation of the gamma function. So, we have

$$\begin{aligned} \frac{\Gamma(r)}{\Gamma\left(r - \frac{1 + b_n}{2\alpha} u + O\left(\frac{1}{n}\right)\right)} &\sim r^{\frac{1 + b_n}{2\alpha} u + O\left(\frac{1}{n}\right)} \\ &\sim \exp\left\{\frac{1 + b_n}{2\alpha} u \ln r\right\} \\ &\sim \exp\left\{\frac{1 + b_n}{2\alpha} u \ln(\rho n - h_n)\right\} \\ &\sim \exp\left\{\frac{1 + b_n}{2\alpha} u (\ln n + \ln \rho + O(h_n/n))\right\} \\ &\sim \exp\left\{\frac{1 + b_n}{2\alpha} u \ln n\right\} \exp\left\{\frac{\ln \rho}{2\alpha} u\right\}. \end{aligned} \tag{8}$$

Following similar steps, we find

$$\frac{\Gamma(j)}{\Gamma(j - \frac{1+b_n}{2\alpha}u + O(\frac{1}{n}))} = \exp\left\{\frac{1+b_n}{2\alpha}u \ln n\right\} \exp\left\{\frac{\ln \alpha}{2\alpha}u\right\} \tag{9}$$

and

$$\frac{\Gamma(j-r+1 - \frac{1+b_n}{2\alpha}u + O(\frac{1}{n}))}{\Gamma(j-r+1)} \sim \exp\left\{-\frac{1+b_n}{2\alpha}u \ln n\right\} \exp\left\{-\frac{\ln(\alpha-\rho)}{2\alpha}u\right\}. \tag{10}$$

The critical term on the right-hand side of each of (8), (9), and (10) is $\ln n$. It appears in these equations twice positively and once negatively. Putting these elements together, we can now assert that

$$\begin{aligned} E[e^{W_{r_n} \frac{u}{n}}] &\sim \exp\left\{\frac{1+b_n}{2\alpha}u \ln n\right\} \exp\left\{\frac{\ln \rho}{2\alpha}u\right\} \exp\left\{\frac{\ln \alpha}{2\alpha}u\right\} \exp\left\{-\frac{\ln(\alpha-\rho)}{2\alpha}u\right\} \\ &\times \Gamma\left(1 - \frac{1}{2\alpha}u\right). \end{aligned}$$

In other words, we have

$$E\left[\exp\left\{\left(W_{r_n} - \frac{1+b_n}{2\alpha}n \ln n\right) \frac{u}{n}\right\}\right] \rightarrow \exp\left\{\frac{u}{2\alpha} \ln\left(\frac{\alpha\rho}{\alpha-\rho}\right)\right\} \Gamma\left(1 - \frac{1}{2\alpha}u\right).$$

The right-hand side is the moment-generating function of a Gumbel random variable with location parameter $(2\alpha)^{-1} \ln(\alpha\rho/(\alpha-\rho))$ and scale parameter $1/(2\alpha)$. By the Lévy continuity theorem [7, p. 172], convergence of moment-generating functions implies convergence in distribution. In other words, we have

$$\frac{W_{r_n} - \frac{(1+b_n)}{2\alpha}n \ln n}{n} \xrightarrow{\mathcal{D}} \text{Gum}\left(\frac{1}{2\alpha} \ln\left(\frac{\alpha\rho}{\alpha-\rho}\right), \frac{1}{2\alpha}\right).$$

This completes a proof of Theorem 3.1.

6.2. Near depletion (up to lower-order perturbation)

Suppose $r_n = \alpha n - h_n$ for perturbation function h_n that is $o(n)$ and $r_n \leq j_n$, so necessarily we have $h_n \geq g_n$. We have to separate two cases: (a) when $h_n - g_n \rightarrow \infty$, and (b) when $h_n - g_n$ remains bounded.

In case (a), the alternative to (10) is then

$$\frac{\Gamma(j-r+1 - \frac{1+b_n}{2\alpha}u + O(\frac{1}{n}))}{\Gamma(j-r+1)} \sim \exp\left\{-\frac{1+b_n}{2\alpha}u \ln(h_n - g_n)\right\}.$$

Note that the $\ln n$ terms in (10) are annulled and $\ln n$ appears only twice positively in $\phi_r(t)$. The alternative statement to Theorem 3.1 is then

$$\frac{W_{r_n} - \frac{1}{\alpha(1-\frac{g_n}{\alpha n})}n \ln n + \frac{1}{2\alpha(1-\frac{g_n}{\alpha n})}n \ln(h_n - g_n)}{n} \xrightarrow{\mathcal{D}} \text{Gum}\left(\frac{\ln \alpha}{\alpha}, \frac{1}{2\alpha}\right).$$

Remark 6.1. Note how the coefficient of $n \ln n$ is doubled as compared to the cases that are far away from depletion.

In case (b), the Stirling approximation is not applicable to the pair of gamma functions in (10), as the argument of both gamma functions remains bounded. Here we have $j_n = \alpha n - g_n$, $r_n = \alpha n - (g_n + x_n)$, with $x_n = O(1)$. So, we have $j_n - r_n = x_n$. There is a large variety in how the term x_n behaves. It might be oscillating, and as we shall see shortly, this does not lead to convergence. We focus on the more interesting case when $x_n = c > 0$ is a constant. So, we have

$$\frac{\Gamma(j - r + 1 - \frac{1+b_n}{2\alpha} u + O(\frac{1}{n}))}{\Gamma(j - r + 1)} \rightarrow \frac{\Gamma(c + 1 - \frac{1}{2\alpha} u)}{\Gamma(c + 1)}.$$

The moment-generating function in this phase is

$$\begin{aligned} \phi_r\left(\frac{u}{n}\right) &= \Gamma(r) \times \frac{\Gamma(1 - \frac{1+b_n}{2\alpha} u + O(\frac{1}{n}))}{\Gamma(r - \frac{1+b_n}{2\alpha} u + O(\frac{1}{n}))} \times \frac{\Gamma(j)}{\Gamma(j - \frac{1+b_n}{2\alpha} u + O(\frac{1}{n}))} \times \frac{\Gamma(c + 1 - \frac{1}{2\alpha} u)}{\Gamma(c + 1)} \\ &\sim \exp\left\{\frac{1 + b_n}{\alpha} u \ln n\right\} \exp\left\{\frac{\ln \alpha}{\alpha} u\right\} \Gamma\left(1 - \frac{1}{2\alpha} u\right) \frac{\Gamma(c + 1 - \frac{1}{2\alpha} u)}{\Gamma(c + 1)}. \end{aligned}$$

The corresponding limit theorem is

$$\frac{W_{r_n} - \frac{1}{\alpha(1-\frac{g_n}{\alpha n})} n \ln n}{n} \xrightarrow{\mathcal{D}} \text{Ojo}\left(c + 1, \frac{1}{2\alpha}\right) + \text{Gum}\left(\frac{\ln \alpha}{\alpha}, \frac{1}{2\alpha}\right), \tag{11}$$

and the Ojo and Gumbel random variables in the limit are independent.

6.3. Complete depletion

In the case of complete depletion, when $r = j$, the variable W_j represents the waiting time until all the potential spreaders are converted into active spreaders. Attention was drawn to this case in the literature of alternative models. In this case, the generalized Gumbel random variable in (11) comes down to an ordinary Gumbel random variable, leading to the convergence relation

$$\frac{W_j - \frac{1}{\alpha(1-\frac{g_n}{\alpha n})} n \ln n}{n} \xrightarrow{\mathcal{D}} \text{Gum}_1\left(\frac{\ln \alpha}{2\alpha}, \frac{1}{2\alpha}\right) + \text{Gum}_2\left(\frac{\ln \alpha}{2\alpha}, \frac{1}{2\alpha}\right),$$

and the two Gumbel random variables Gum_1 and Gum_2 are independent. Only for æsthetic reasons (to symmetrize the two Gumbel random variables) do we split the term $e^{\frac{u}{\alpha} \ln \alpha}$ in the limiting moment-generating function into two equal parts; one part is subsumed as a location parameter in Gum_1 , the other is taken as a location parameter in Gum_2 .

Acknowledgements

The author is indebted to Sheldon Ross for insightful discussions.

References

[1] BELEN, S., KROPAT, E. AND WEBER, G. (2009). On the classical Maki–Thompson rumour model in continuous time. *Cent. Eur. J. Oper. Res.* **19**, 1–17.
 [2] DALEY, D. AND KENDALL, D. (1965). Stochastic rumours. *J. Inst. Math. Appl.* **1**, 42–55.
 [3] DOUMAS, A. AND PAPANICOLAOU, V. (2012). The coupon collector’s problem revisited: Asymptotics of the variance. *Adv. Appl. Prob.* **44**, 166–195.
 [4] FERRANTE, M. AND SALTALAMACCHIA, M. (2014). The coupon collector’s problem. *MATerials MATemàtics* **2**, 1887–1907.

- [5] FRAUENTHAL, J. (1980). *Mathematical Modeling in Epidemiology*. Springer, Berlin.
- [6] JUNIOR, V., MACHADO, F. AND ZULUAGA, M. (2011). Rumor processes on \mathbb{N} . *J. Appl. Prob.* **48**, 624–636.
- [7] KARR, A. (1993). *Probability*. Springer, New York.
- [8] MAKI, D. AND THOMSON, M. (1973). *Mathematical Models and Applications*. Prentice-Hall, Englewood Cliffs, NJ.
- [9] MOLCHANOV, S. AND WHITMEYER, M. (2010). Two Markov models of the spread of rumors. *J. Math. Sociol.* **34**, 157–166.
- [10] OJO, M. (2001). Some relationships between the generalized Gumbel and other distributions. *Kragujevac J. Math.* **23**, 101–106.
- [11] PITTEL, B. (1987). On spreading a rumor. *SIAM J. Appl. Math.* **47**, 213–223.
- [12] PITTEL, B. (1990). On a Daley–Kendall model of random rumours. *J. Appl. Prob.* **27**, 14–27.
- [13] SUDBURY, A. (1985). The proportion of the population never hearing a rumour. *J. Appl. Prob.* **22**, 443–446.
- [14] SVENSSON, A. (1993). On the duration of a Maki–Thompson epidemic. *Math. Biosci.* **117**, 211–220.
- [15] WATSON, R. (1987). On the size of rumour. *Stoch. Proc. Appl.* **27**, 141–149.
- [16] WHITTAKER, E. AND WATSON, G. (1990). *A Course in Modern Analysis*, 4th edn. Cambridge University Press.