Modelling Twitter Trends

0803155
(Dated: March 30, 2012)

This project researches and attempts to mathematically model the spread of popular topics on the popular social networking site Twitter. This modelling is based on the Susceptible Infected Recovered (SIR) epidemiological model, drawing parallels between the spread of infections and the spread of topics on Twitter. The SIR model has been modified to reflect certain mechanisms of Twitter. It has been discovered that introducing reinfection, to reflect users spontaneously rejoining the trend, accurately models the long tail often found in graphs of the evolution of phrase popularity. The effect of the scale-free nature of the Twitter network on the spread of a topic has been considered. This reveals that phrases do not need to be innately infectious in order to trend and the total size of an epidemic will be smaller on a scale-free network than for an equivalent homogeneous network.

I. INTRODUCTION

Twitter is a microblogging website that lets users share their thoughts in posts of 140 characters or less, and read the posts, or 'tweets' as they are called, of users they wish to follow. Recently its significance has grown considerably due to the impact it is having on information dissemination in multiple spheres of society. It is becoming increasingly difficult for authorities to control the flow of information as this process is democratised by Twitter and similar digital social networks. The power of this was clearly demonstrated by the planning of protests in Iran [1] and riot clean-up operations in London [2] both orchestrated almost entirely via Twitter. The debacle of super injunctions in the UK and their impotence in the face of this now mighty networking tool also shows the powerful effect Twitter is having on our society [3].

This project researches one particular aspect of Twitter, trends. When a significant number of tweets within a short period of time contain a particular phrase the phrase is said to trend on Twitter. These phrases are often in the form of a hashtag (there is a glossary of relevant terms in Appendix A) which is a form of social annotation for a topic, event or meme. There is a constant list of currently trending topics on the Twitter homepage. This list gives an insight into what people are currently tweeting about and as such contains a wealth of information on what is immediate and popular.

As the influence of social networks grows, it is becoming increasingly important to understand how information disseminates within these networks. I have attempted to model these trends and thus discover more in depth information about them.

This paper will be structured as follows. First the data that has been obtained on Twitter trends will be discussed. Then basic disease models, in particular the SIR model, will be explained and how they may be used to model this data. Several modifications to the SIR model will be introduced that reflect certain mechanisms of trend propagation on Twitter. These will be examined and several of them compared with trend data that has been obtained. The network structure of Twitter and the implications of introducing network considerations into a model will then be discussed. Finally the ways this project can be developed further will be explored.

II. DATA

To check the validity of the models in this project, data on trends was collected. It was intended that a large data sample of tweets would be collected to examine. However it was found that it is necessary to be whitelisted by Twitter to gain full access to the Twitter API and to have a dedicated system for collecting data. These were both beyond the scope of this project. The limited access that was obtainable was insufficient to gather more than small samples. These, although representative, are not large enough to perform truly meaningful statistical analysis.

For this project we are only concerned with how the number of tweets containing a trending phrase changes in time. Information such as who tweeted a post or what they actually said is not needed. Thus it was unnecessary to collect complete data on every single tweet containing a trending phrase. A website called Trendistic [4] was used to gather data. This website can search for a phrase or hashtag and produce a graph showing the percentage of total tweets that contained that phrase over time. Since it was not possible to collect a large sample of tweets, Twitter's own trending algorithm, which decides what topics 'trend', was relied upon to discover trends. Trendistic was used to observe a large number of trends and collect data on 35 in particular, 10 of which feature in this report. Details of these featured trends can be found in Appendix B.

Trendistic cannot guarantee that it finds every single tweet containing a search term. If the percentage of tweets containing the trending phrase was significant the results are accurate enough for the purposes of this project. Thus only trends which constituted at least 0.1% of all tweets on Twitter in any one hour have been considered.

The evolution of the popularity of phrases within tweets has been discretised into three classes in a very recent paper by Goncalves *et al.* [5]. This paper, which had many similar aims to this project, concluded that the dynamic behaviour of hashtags follows three behaviour profiles: continuous activity, periodic activity and isolated peak activity. The first class consists of phrases that are continuously found in tweets, such as standard English expressions and common sentiments. (It is interesting to note that the popstar 'Justin Bieber' is consistently found in around 0.1% of all tweets and Twitter specifically tweaked their algorithm so he did not trend constantly!) Periodic activity follows regular cycles such as 'church' peaking every Sunday and #followfriday, which is a Twitter tradition of recommending users to follow every Friday. The last class of phrase popularity is characterised by activity concentrated around a single isolated peak. These are phrases which grow in popularity for a brief period before losing their novelty and disappearing. These are the phrases that Twitter calls trending topics of which this project is concerned.

The number of users on Twitter at any one time massively varies based on circadian cycles and thus so does the number of tweets. This means that knowing the exact number of tweets posted in an hour that contain a particular hashtag or phrase may not be representative of its popularity depending on what time of day the particular activity occured. Therefore we will consider the percentage of total tweets within an hour that contain a trending phrase. This removes the effect of daily cycles and allows the success of different trends to be fairly compared regardless of when they trended. For simplicity it has been assumed in this project's models that the total number of tweets in any hour remains constant.

A time scale of hours has been used as Trendistic only produces values for hourly intervals which in turn is due to limitations of the Twitter API. This granularity is not as precise as would have been desired since many topics only actually trend for the order of 20-40 minutes [6] but is sufficient to give the general behaviour of a trend.

III. THE SIR MODEL

When something quickly becomes popular on the internet it is often said to have gone 'viral'. It is intuitive to think of ideas or memes spreading like a disease. Instead of passing on infections, people on the internet pass on comments, videos, links, quotes *etc.* Parallels have been drawn previously between epidemic dynamics and spreading phenomena on social networks [7]. Infection models have also been thoroughly studied; thus it is natural to base an attempt to model Twitter trends on disease models.

One of the most fundamental disease models is Kermack and McKendrick's Susceptible Infected Recovered or SIR model [8]. The simple epidemic case, where the total population is assumed to be constant, will be used. This is used to model diseases that spread much more rapidly than population changes and so birth and death rates are considered to be zero. The SIR model divides an entire population into three classes: Susceptible, Infected and Recovered.

Initially the population is considered entirely susceptible. A very small number of infected individuals are introduced into this completely susceptible population. These infected individuals interact with susceptible individuals and upon each interaction have a fixed chance of infecting them. When individuals become infected they spontaneously recover at a set rate. Once they have recovered they gain immunity from the infection and remain recovered indefinitely.

The model is defined by the following deterministic mean field differential equations:

$$\frac{dS}{dt} = -\beta SI \tag{1}$$

$$\frac{dI}{dt} = \beta SI - \gamma I \tag{2}$$

$$\frac{dR}{dt} = \gamma I \tag{3}$$

where S, I and R are the proportions of the total population of the Susceptible, Infected and Recovered classes respectively. Note that $S + I + R = 1$ as they are proportions. The transmission rate between infected and susceptible individuals is denoted by β and γ is the rate at which an infected individual recovers.

Eq. (1) states that the size of the susceptible class changes at a rate proportional to the product of the proportions of the susceptible and infected classes. This assumes random homogeneous mixing which is equivalent to the Law of Mass Action. This states that the rate of contact between two groups is proportional to the size of each group concerned [9]. In other words every individual has an equal random chance of interacting with any other individual within the population. Thus the chance of a susceptible individual interacting with an infected individual is directly proportional to their total sizes in the population. Eq. (1) models susceptible individuals becoming 'infected' by interacting with a random infected individual and contracting the infection. The transmission rate between a pair of infected and susceptible individuals is β. Thus these individuals leave the susceptible class.

Eq. (2) states that the individuals leaving the susceptible class immediately join the infected class. Infected individuals spontaneously recover at rate γ and leave the infected class.

Eq. (3) states that the individuals leaving the infected class immediately join the recovered class where they remain indefinitely.

An important quantity in this model is the basic reproductive ratio

$$R_0 = \frac{\beta}{\gamma}$$

This is the average number of secondary cases produced by an average infectious individual in a totally susceptible population. This value dictates the final size of the epidemic and also the peak size of the infected class in an epidemic. It is clear to see that if $R_0 > 1$ the infection will spread and become an epidemic and if $R_0 \leq 1$ the infection will decrease and die out. This is also called the epidemic threshold.

The SIR system is a fixed point when the size of the infection class is zero as then clearly all three differential equations are zero. The entirely susceptible stationary point

$$S = 1, \quad I = 0, \quad R = 0$$

is unstable for $R_0 > 1$ (since this causes an epidemic) and stable for $R_0 \leq 1$. There clearly cannot be a stationary point with non-negative infection. When the system is in the state

$$S = 1 - R, \quad I = 0, \quad R > 0$$

it is also in equilibrium for any $0 < R < 1$.

Since the susceptible class can only decrease, eventually the amount of new infection must also decrease as there are less individuals to infect in the population. This means we can establish the finite final size of an epidemic which is the total number of individuals who ever became infected. Since every infected individual must eventually recover this is the size of the recovered class as time tends to infinity. So at infinity the system must be at a stationary point.

If we divide Eq. (1) by Eq. (3) and rearrange we get

$$\frac{dS}{dt} = -R_0 S \frac{dR}{dt}$$

which implies

$$S(t) = S(0)e^{-R_0(R(t)-R(0))}$$

where $S(0)$ and $R(0)$ are the initial proportions of the susceptible and recovered classes and R_0 is the basic reproductive ratio. Now since

$$S(t) + I(t) + R(t) = 1 \text{ and } R(0) = 0$$

we get

$$S(0)e^{-R_0 R(t)} + I(t) + R(t) = 1$$

The maximum proportion of the population the epidemic can ever infect is the total population, although it is more likely to be smaller. So eventually we must reach an equilibrium where $I = 0$ and there can be no more infection. So if we rearrange and send $t \to \infty$ we get

$$R_\infty = 1 - S(0)e^{-R_0 R_\infty}$$

This value can be calculated and we can see that R_∞ approaches 1 as R_0 becomes significantly larger than one.

$$S = 1 - R_\infty, \quad I = 0, \quad R = R_\infty$$

is an important stable stationary point of this system.

The SIR system models infections where lifelong immunity is gained after infection. It assumes transmission is frequency dependent, which has been experimentally shown to accurately model human diseases such as measles [10]. The SIR model has also been extensively studied and extended so there is a wealth of knowledge to call upon. Is this model suitable for modelling Twitter trends however?

For endogenous trends which are originated by a few individuals on Twitter and do not have exogenous forcing, this model makes intuitive sense. It would be very difficult to model powerful exogenous forcing by, for example, mass media and news stories as the effect of these can massively vary. There is also no clear practical way to measure such forcing so this will not be considered here.

The biggest assumption of the above SIR model is that an interaction between any two particular individuals is equally likely as any other, which ignores network structure. This will be discussed later. Since disease transmission normally occurs due to social contact it is reasonable to initially assume that frequency dependent transmission is representative of virtual social contacts as well.

All the online twitter users can be thought of as our population who are initially susceptible. A user or several users start a trend or ‘infection’ by tweeting a phrase or hashtag. Fellow users who are following their posts see one of these tweets and have a chance of becoming infected by passing on the phrase. This can be done by passing on the tweet directly to their own followers, which is called retweeting, or write a tweet themselves containing the phrase. Retweets make up approximately 31% of the tweets of trending topics [6] however we will not differentiate between these and original tweets. Thus by tweeting the phrase they spread the ‘infection’ to other users who then themselves spread it by tweeting it.

If the ‘infection’ becomes a large enough epidemic it becomes a trending topic. We consider users who tweet or retweet a trend phrase as our infected class. They remain infected while they continue to tweet using the phrase until they become ‘recovered’ *i.e.* when they stop tweeting about it. It will be assumed that each online user tweets an equal amount an hour; so the number of tweets in an hour is directly proportional to the number of users. Thus the size of the infected class, $I(t)$, at hourly time intervals, t, is equivalent to the number of tweets containing the concerned phrase within the previous hour.

The worldwide, and sometimes national, trending topics published by Twitter were regularly recorded. The evolution of their popularity was then observed using Trendistic to see if the epidemic behaviour predicted by the SIR model is exhibited on Twitter. Fig. 1 shows four trends observed that exhibited the single peak behaviour very similar to that seen in SIR models. This shows SIR model behaviour is exhibited on Twitter.

Next a SIR model was fitted to an actual trend. Matlab’s ODE solver was used to solve the SIR equations.

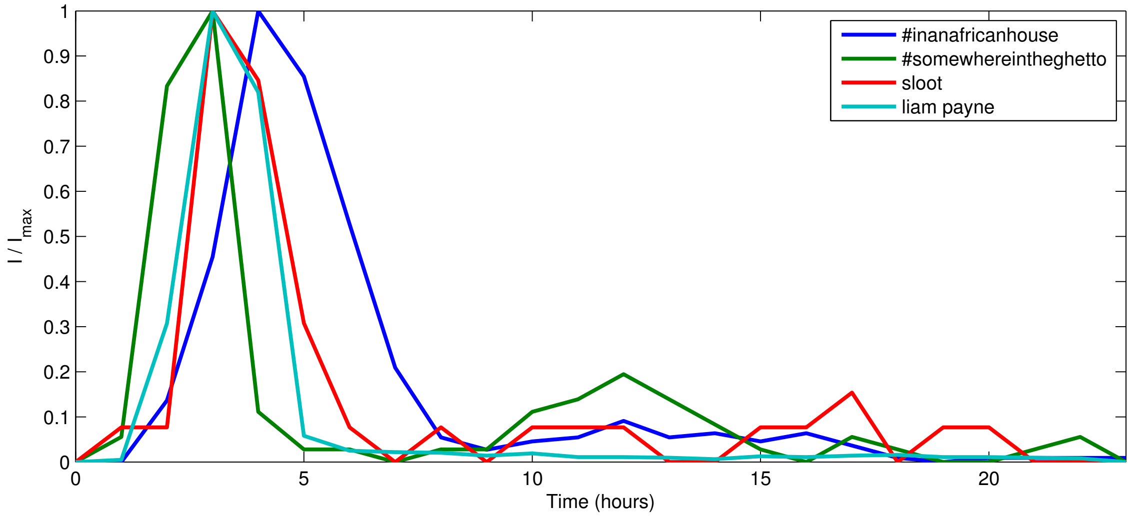

FIG. 1: Scaled popularity activity for four Twitter trends that exhibit single peak behaviour.

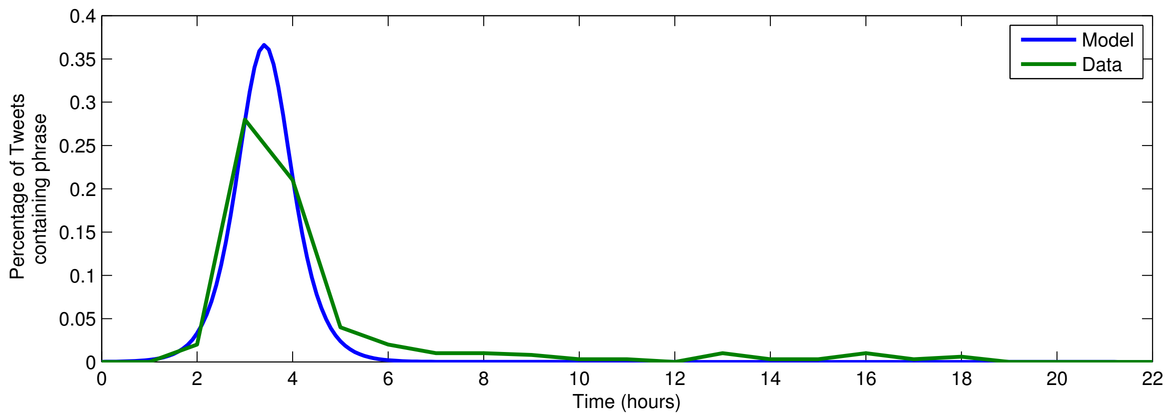

FIG. 2: Data collected for trend #sotellmewhy displayed in green alongside an approximate SIR model in blue.

These equations were applied to a population of 100,000 to achieve necessary accuracy. On 2nd January 2012 the hashtag '#sotellmewhy' trended on Twitter. Data on this trend was collected using Trendicity. Matlab was then used to find the values for β, γ and initial infected, I_0, that produce the closest approximation to the data. This was done by writing a function, $I(\beta, \gamma, I_0, t)$, that, given β, γ and I_0, uses the ODE solver to give the size of the infected class at time t. Next Matlab's *fminsearch* function was used to find values of β, γ and I_0 that minimise

$$\left(\sum_t (I(\beta, \gamma, I_0, t) - D(t))^2 \right)^{\frac{1}{2}}$$

where $D(t)$ is the data at time t.

The original data and model approximation are displayed in Fig. 2. The initial conditions of 100,000 population, 0.1591 initial infected, $\beta = 31.9229$ and $\gamma = 29.2331$ were used to produce the approximation in the figure shown in blue.

Visually it is clear that the simple endemic SIR model can be used to model this particular trend. This was not the only trend that was observed to display this simple behaviour, see Fig. 1, however such trends were a minority. It was observed that most trends exhibit more complicated behaviour. More data is required to discern precisely how prevalent this simple endemic behaviour is.

IV. SIR WITH SYSTEMIC FORCING: THE EFFECT OF THE TWITTER TRENDING TOPICS LIST

One important consideration for modelling endogenous trends is the list of trending topics that appears on the Twitter homepage. Users learn of endogenous trends not only from the users they are following but may also dis-

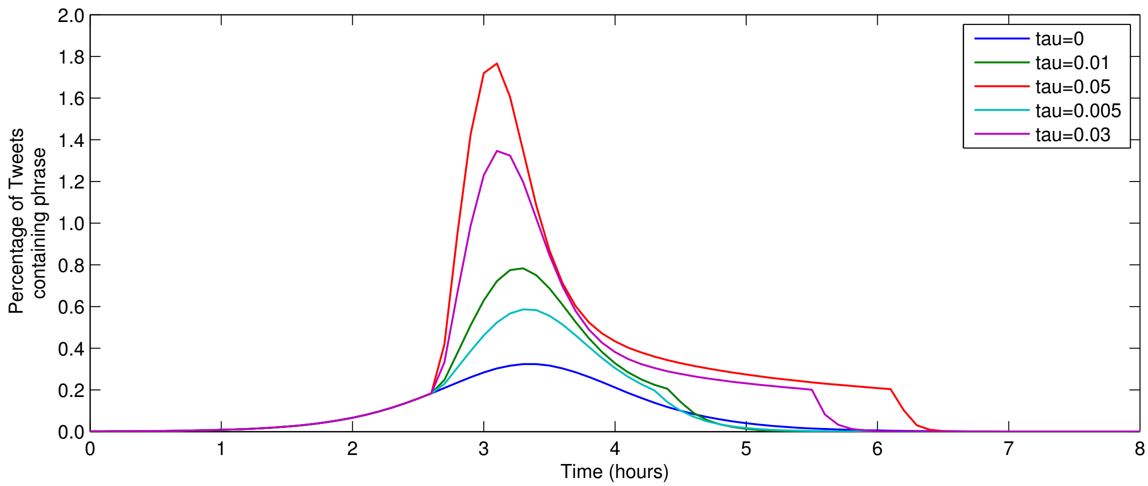

FIG. 3: SIR with Forcing model for $\beta = 31.9229$, $\gamma = 29.2331$, $I_0 = 0.1591$, $\theta = 0.2$ and a range of values for τ.

cover them from this list and thus become 'infected'.

Twitter's algorithm for deciding whether something is considered a trend is secret, presumably to avoid being abused. It is known that it depends on both the number of users that tweet a phrase, the number of tweets and the duration of time for which it is tweeted about. Trends must also compete with other trends to be listed on the trending topics table.

The selection algorithm has been approximated by assuming it only requires a fixed proportion of tweets within an hour to contain a phrase for it to be listed as a trend. It has also been assumed that the phrase remains on the list for as long it is above this value.

To model this, the SIR equations were modified:

$$\frac{dS}{dt} = -\beta SI - \tau S\chi(I,\theta) \tag{4}$$

$$\frac{dI}{dt} = \beta SI + \tau S\chi(I,\theta) - \gamma I \tag{5}$$

$$\frac{dR}{dt} = \gamma I \tag{6}$$

where θ is the threshold proportion required for a trend to appear on the list of trending topics. $\chi(I,\theta)$ in Eq. (4) and Eq. (5) is an approximate indicator function defined by:

$$\chi(I,\theta) = \left(\frac{\tanh(\alpha(I-\theta))}{2} + \frac{1}{2}\right)$$

For $I \le \theta$, $\chi(I,\theta) = 0$, and $I >> \theta$, $\chi(I,\theta) = 1$. We introduce α as a coefficient large enough such that the function is steep around θ. So $\chi(I,\theta) = 1$ for values of I just above θ. The value $\alpha = 100,000$ was used for calculations.

This function has been used instead of a discontinuous indicator function as its continuity prevents problems with Matlab's ODE solver. It means for sufficiently large infection there is an additional force of infection, defined by τ, proportional to the size of the susceptible class.

When the tweets containing a hashtag or phrase constitute more than θ of all tweets within an hour, this indicates that Twitter considers it a trend and lists it as a trending topic. While it is a trending topic, susceptible individuals spontaneously become infected by seeing it. This happens at a rate proportional to the size of the susceptible class.

This system, like the simple SIR model, is in equilibrium when there is zero infection. Just like before, there is a stationary point when the entire population is susceptible . This system will give a different larger value for R_∞ which must be measured numerically.

Fig. 3 shows the effect of different levels of forcing. The value $\theta = 0.2\%$ has been used as an approximation taken from unverified data found online [11]. For $\tau = 0$ the model is identical to the non-forcing SIR model from before. The blue line in Fig. 3 is identical to the blue line in Fig. 2 when there is no forcing in the model. The behaviour of the infection is identical to the non-forcing model until the infection reaches θ and it will be completely identical if the infection never reaches θ.

As can be observed from the graph, upon reaching θ infection increases noticeably and, as is expected, achieves increasing levels of infection for higher values of τ. The peak of infection also occurs earlier and the epidemic starts diminishing sooner. For low values of τ the size of infection can actually be less than for the non-forcing model towards the end of the epidemic. This can be seen in Fig. 3 for $\tau = 0.005$ and $\tau = 0.01$.

After the peak, the infection decreases until it again passes the θ threshold, upon which it decreases rapidly. This behaviour creates a 'long tail'. Varying θ only changes the threshold that an epidemic must reach before having the behaviour described above. A reliable value for θ was not found so this model has not been used to fit data.

This behaviour would be expected to be exhibited by all the trends in the collected data since all these trends were *de facto* listed as a trending topic. However since data has not been collected for phrases that did not trend we have no comparison with which to establish values for τ. We would expect however for it to be related to β. The 'infectiousness' of a trending topic has been shown to depend more on the inherent appeal of the topic rather than other factors such as the attributes of 'infected' individuals [6]. Thus how likely a user is to start tweeting a phrase is more dependant on the content than where they discovered it. The 'infectiousness' of a phrase however is a vague concept, one that would be very difficult to quantify.

V. SIR MODEL WITH TIME DEPENDENT REINFECTION: SPONTANEOUSLY TWEETING ON A TREND AGAIN LATER

It has been assumed thus far that once a user stops tweeting on a trend, they do not restart tweeting about it at some later point. Also it has been assumed that all users 'recover' at the same rate. It is quite likely that some users will continue to tweet on a trend for an extended length of time, especially if they enter a discussion on the topic with several other users; others will only tweet once or twice. This could be modelled by introducing different subclasses of the infected class that have different recovery rates. However this introduces many new variables that cannot easily be estimated or measured. This concept will be modelled by having some recovered individuals spontaneously become reinfected. This can be done by amending the SIR equations:

$$\frac{dS}{dt} = -\beta SI \tag{7}$$

$$\frac{dI}{dt} = \beta SI - \gamma I + \zeta R\chi^{'}(t,T) \tag{8}$$

$$\frac{dR}{dt} = \gamma I - \zeta R\chi^{'}(t,T) \tag{9}$$

Here the time dependent term, $\chi^{'}(t,T)$, has been introduced in Eq. (8) and Eq. (9) defined by

$$\chi^{'}(t,T) = \left(\frac{1}{2} - \frac{\tanh(\alpha(t-T))}{2}\right)$$

This is another continuous approximate indicator function. So while time t is less than some value T a proportion of the recovered class, set by the ζ term, will become infected again. This models users who tweet briefly on a trend and then spontaneously decide to tweet again about it later. The period of time for which 'recovered' users may become reinfected has been averaged as T.

For finite T the system will eventually become the simple epidemic SIR system and thus it has the same stationary points, when infection is zero.

When very large values of T are considered, the system tends to an equilibrium. While $t < T$ there are always some recovered individuals becoming reinfected so there will always be infected individuals in the population. So, since the number of susceptibles can only decrease, eventually all of the susceptible class must become infected for a large enough T. If this point is reached the system will tend to an equilibrium where the sizes of the infected and recovered classes is constant. By solving Eq. (8) and Eq. (9) for

$$S = 0 \quad \text{and} \quad \chi^{'}(t,T) = 1$$

we find this equilibrium state is

$$S = 0, \quad I = \frac{\zeta}{\gamma+\zeta}, \quad R = \frac{\gamma}{\gamma+\zeta}$$

This equilibrium will only remain while $t < T$ however.

Initially the recovered class is very small so the epidemic behaves very similarly to the simple epidemic. However as time continues, the recovered class grows and the reinfection of recovered individuals has an increasing forcing effect on the infected class. Thus we would expect to see an infection peak, as in the simple epidemic case, and then for the infection to tend towards the above equilibrium while $t < T$. How fast the system tends towards equilibrium depends how fast all of the remaining susceptibles become infected. After time T the epidemic behaves once again like the simple epidemic as there is no reinfection, so the epidemic dies out and the sizes of the susceptible and recovered classes become constant. If T is large enough these will be zero and the total population respectively. Hence the size of T can have significant impact on the value of final epidemic size.

As can be seen in Fig. 4 this behaviour is exhibited by trends on Twitter. This model was fitted to a trend. On 4th January 2012 the hashtag '#ISecretlyLove' trended. Data was collected on it using Trendistic and values found to produce Fig. 5. A very similar Matlab technique as with the simple SIR model was used to establish the best values for β, γ, ζ, T and I_0. Due to the coarse granularity of the data multiple sets of values were found using *fminsearch* that equally minimised

$$\left(\sum_t (I(\beta,\gamma,I_0,\zeta,T,t) - D(t))^2\right)^{\frac{1}{2}}$$

The values given here were chosen as they produced the median approximation.

A population of 100,000 was used, as with all numerical analysis of the models in this project. The following values were calculated: $\beta = 24.5329$, $\gamma = 22.4415$, $\zeta = 0.0051$, $T = 13.3882$ and $I_0 = 1.0209$. These produced the approximation in Fig. 5 shown in blue.

Adding this ζ term can model trends which exhibit this long tail as shown in Fig. 4. These trends were prevalent in the observations of this project, however these tails are not always smooth and often fluctuate. It is not

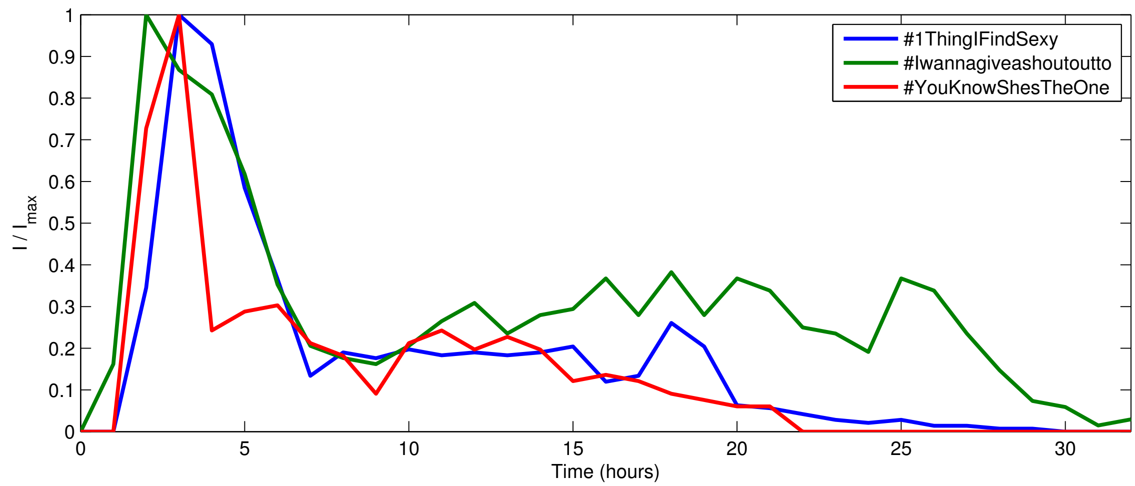

FIG. 4: Scaled popularity activity for three Twitter trends that exhibit single peak behaviour with a long tail.

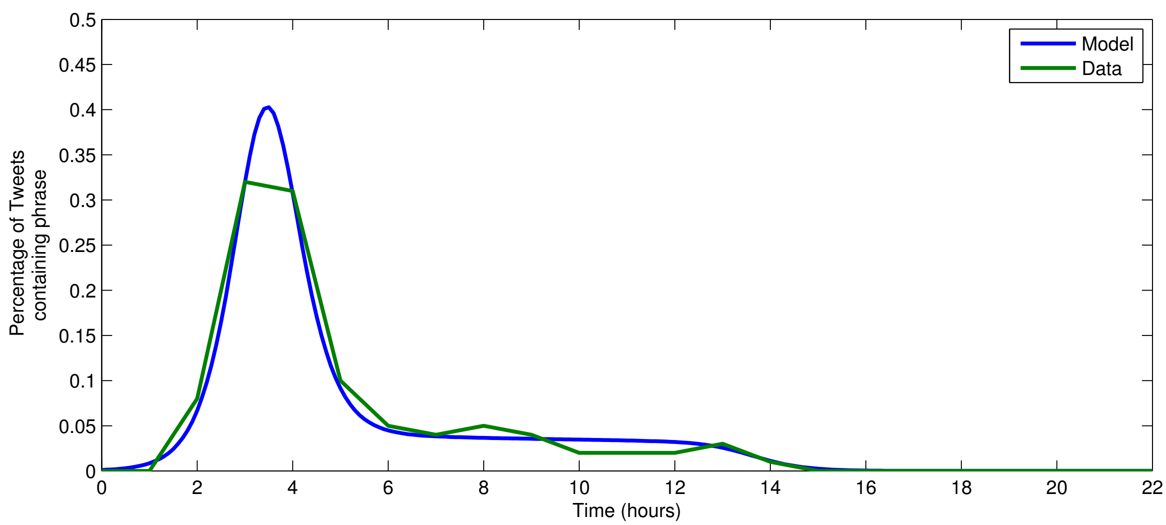

FIG. 5: Data for trend #ISecretlyLove displayed in green alongside an approximate SIR model with time dependent reinfection in blue.

easy to discern whether this is due to stochastic effects or other forces at work. I expect it is most likely it is a combination of both.

VI. SEIR MODEL: DELAY IN TWEETING

It has been assumed thus far that when a user sees someone they are following post on a trend and becomes 'infected', they immediately start tweeting on the trend. In reality many users are not constantly checking their Twitter feed. When they do they may see a post from a few hours before and belatedly join in with the trend. This adds an element of delay to some tweets. Thus we will continue using ideas from epidemiology and introduce an 'Exposed' class denoted E. This addition to the SIR model (which is not the same as the standard SEIR model) gives the following equations:

$$\frac{dS}{dt} = -\beta SI \tag{10}$$

$$\frac{dE}{dt} = \epsilon\beta SI - \delta E \tag{11}$$

$$\frac{dI}{dt} = (1-\epsilon)\beta SI + \delta E - \gamma I \tag{12}$$

$$\frac{dR}{dt} = \gamma I \tag{13}$$

In Eq. 11 two new variables, ϵ and δ are introduced. The proportion of susceptibles that do not immediately tweet when a user they are following tweets on a trend,

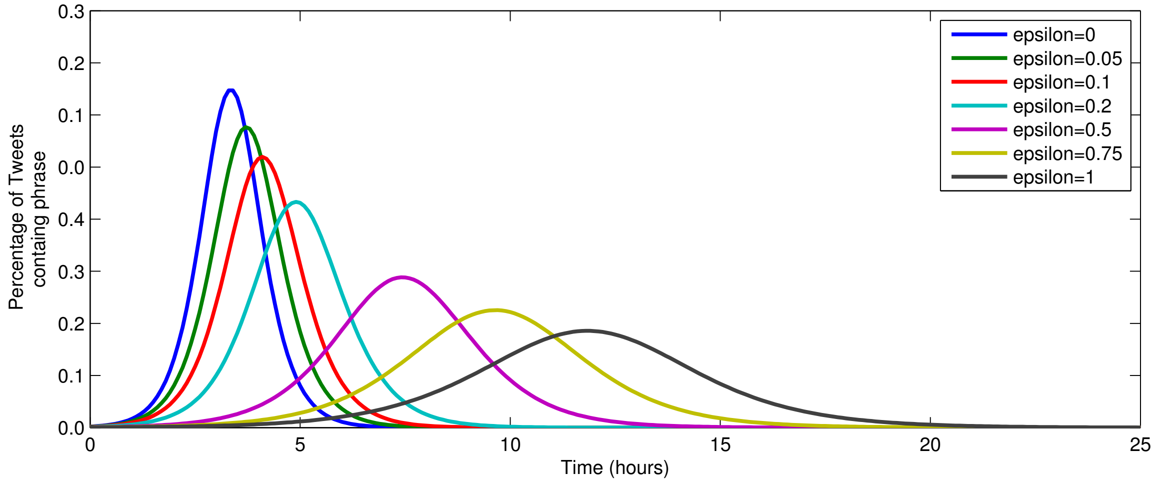

FIG. 6: SEIR model for $\beta = 27.12$, $\gamma = 24.97$, $I_0 = 1$, $\delta = 10$ and varying values of ϵ.

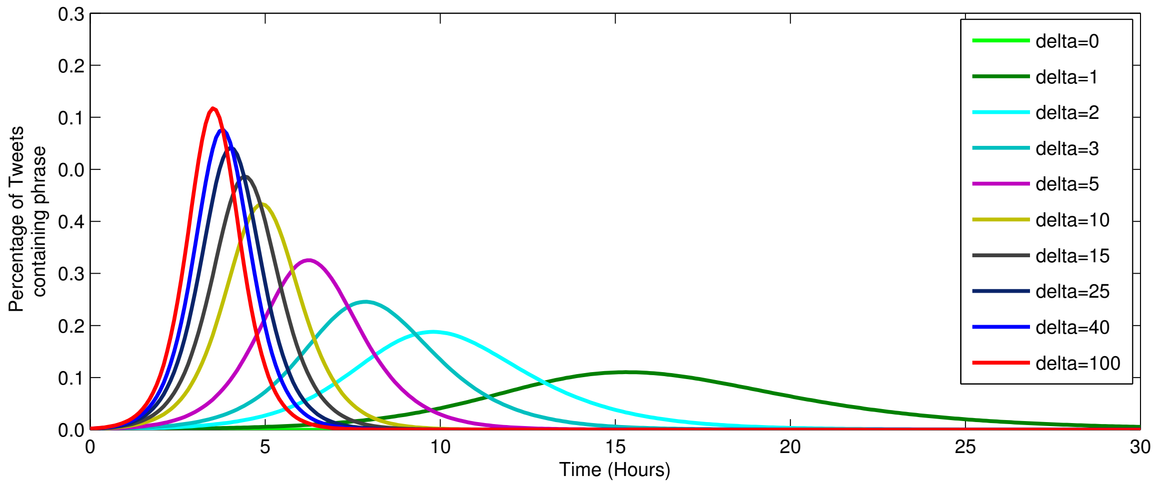

FIG. 7: SEIR model for $\beta = 27.12$, $\gamma = 24.97$, $I_0 = 1$, $\epsilon = 0.2$ and varying values of δ.

but become ‘exposed’, is denoted by ϵ. They later become ‘infected’ by tweeting the trending phrase. This is modelled by the δE term. In Eq. 12 we have the $(1-\epsilon)$ proportion of the susceptibles who, when exposed to infection, immediately start tweeting, and a δE term representing the exposed individuals who have become belatedly infected. Once again individuals can only move linearly through the classes. So if an individual becomes ‘exposed’ they must eventually become infected, then recovered and then remain recovered indefinitely.

Clearly if $\epsilon = 0$ this model is the same as the simple epidemic case. The exposed class behaves very similarly to the infected class in the simple epidemic model but with a transmission rate of $\epsilon\beta$ and recovery rate of δ. Thus we would expect the exposed class to exhibit single peak behaviour.

Fig. 6 shows the effect of varying ϵ for a fixed δ. The $\epsilon = 0$ case is the same as the basic SIR with $\beta = 27.12$, $\gamma = 24.97$ and 1 initial infected. It can be seen that as ϵ increases the graph stretches horizontally. This is intuitive as we would expect R_∞ to remain similar, but the infected activity to be more dispersed.

Fig. 7 shows the model for varying values of δ. We can see it has the opposite effect to ϵ. As δ increases the system becomes more like the SIR case as individuals are exposed for increasingly shorter periods. Note that when $\delta = 0$ there is no epidemic. This is due to the system becoming almost a basic SIR system with transmission rate $\epsilon\beta$. This gives

$$R_0 = \frac{\epsilon\beta}{\gamma} < 1$$

which means the disease dies out and there is no epidemic.

This system has again an infection free equilibrium. The stationary points will be the same as for the simple

SIR model, however the total number of individuals ever infected, R_∞, will depend on ϵ and δ so the post-epidemic stationary point will have different values. There is no endemic equilibrium since the susceptible class is not replenished and there are no loops in the system.

Thus introducing this addition to the model alters the life span of the epidemic and changes the reproductive ratio, but does not affect the topology of the epidemic. Since it was not possible to glean data about the delays between a user tweeting a hashtag and their followers tweeting it, values cannot be established for this model. However, this delay is an important consideration in any such model.

VII. SIR WITH BIRTHS AND DEATHS: ATTRITION OF ONLINE USERS

So far, the population of Twitter users online has been considered to be constant. What has not been factored in is that even if the total number remains constant, the identity of the users changes. This change of users who are online is called attrition. The concepts of births and deaths from disease models can be introduced to the model but instead model users going online or going offline. We will assume that users who are 'online' are actually active on Twitter. The population can be kept constant by having the same number of births as deaths. In the case of our model this means the birth rate is $d(S+R)$, where d is the death rate or going offline rate. This addition can be modelled by the following equations:

$$\frac{dS}{dt} = dR - \beta SI \tag{14}$$

$$\frac{dI}{dt} = \beta SI - \gamma I \tag{15}$$

$$\frac{dR}{dt} = \gamma I - dR \tag{16}$$

Infected individuals do not go offline since they are currently tweeting on the trending topic so must be active. This models the attrition of users online and changes the dynamics of the system since the susceptible class now gets replenished.

We still have the infection free equilibrium and

$$R_0 = \frac{\beta}{\gamma}$$

There is now also a new fixed point where the infection is endemic. If we set $\frac{dI}{dt} = 0$ we get the susceptible stationary point:

$$S^* = \frac{\gamma}{\beta}$$

and from $\frac{dS}{dt} = 0$, using S^* from above we get the infected stationary point:

$$I^* = \frac{d(\beta - \gamma)}{\beta(\gamma + d)}$$

Looking at the Jacobian of this fixed point we get:

$$J = \begin{pmatrix} -\frac{d(\beta+d)}{\gamma+d} & -(\gamma+d) \\ \frac{d(\beta-\gamma)}{\gamma+d} & 0 \end{pmatrix}$$

which has the eigenvalues:

$$\lambda = \frac{-d(\beta+d) \pm \sqrt{d^2(\beta+d)^2 - 4d(\beta-\gamma)(\gamma+d)^2}}{2(\gamma+d)}$$

Assuming d is small compared to the other terms this approximates to:

$$\lambda = -\frac{dR_0}{2} \pm 2i\sqrt{d(\beta-\gamma)\gamma^2}$$

which shows that if $R_0 > 1$ this endemic fixed point is stable and is converged to with damped oscillations. These oscillations are shown in Fig. 8.

Multiple peaks have been observed in the trends in the collected data; however most do not display clearly oscillatory behaviour. Fig. 9 shows a trend that does appear to display oscillatory behaviour but only temporarily. The trends that have been considered do not display continuous activity so infection cannot become endemic. Therefore although attrition of users must be a factor in propagation of topics, it must not be significant and another mechanism must be causing these observed multiple peaks.

VIII. NETWORK STRUCTURE CONSIDERATIONS

The models discussed assume random mixing, where any individual has an equally likely chance of interacting with any and every other individual in the entire population. This assumption has been shown to be valid for modelling certain diseases such as measles [10]. However the network structure of Twitter is considerably different from face-to-face social interaction networks.

So what kind of network structure does Twitter have? Twitter is a directed network where we can consider the users to be nodes. A directed connection between two nodes represents a user following another user's tweets. Statistical analysis has shown the Twitter network displays a scale-free structure [12].

Scale-free networks are defined by the degree of nodes (or number of followers a user has) following a power-law distribution. In other words

$$P(k) \sim k^{-\sigma}$$

where $P(k)$ is the proportion of nodes in the network of degree k. Normally $2 < \sigma < 3$, although it is occasionally outside these bounds. A scale-free network can be generated from an initial network by adding nodes one at a time that preferentially connect to nodes of high degree. This can be achieved by using the Barabási-Albert or

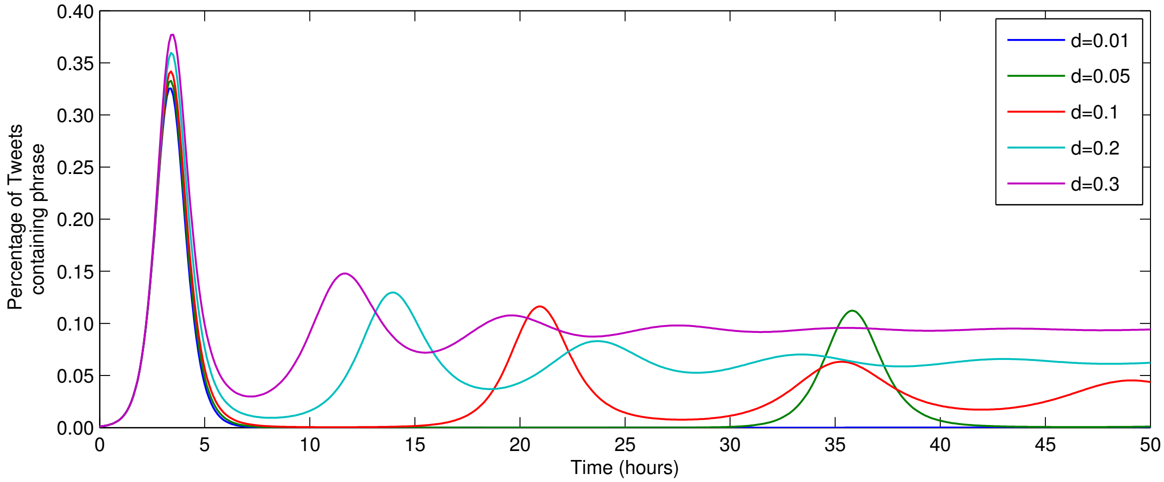

FIG. 8: SIR with attrition model for $\beta = 27.12$, $\gamma = 24.97$, $I_0 = 1$ and varying values of d.

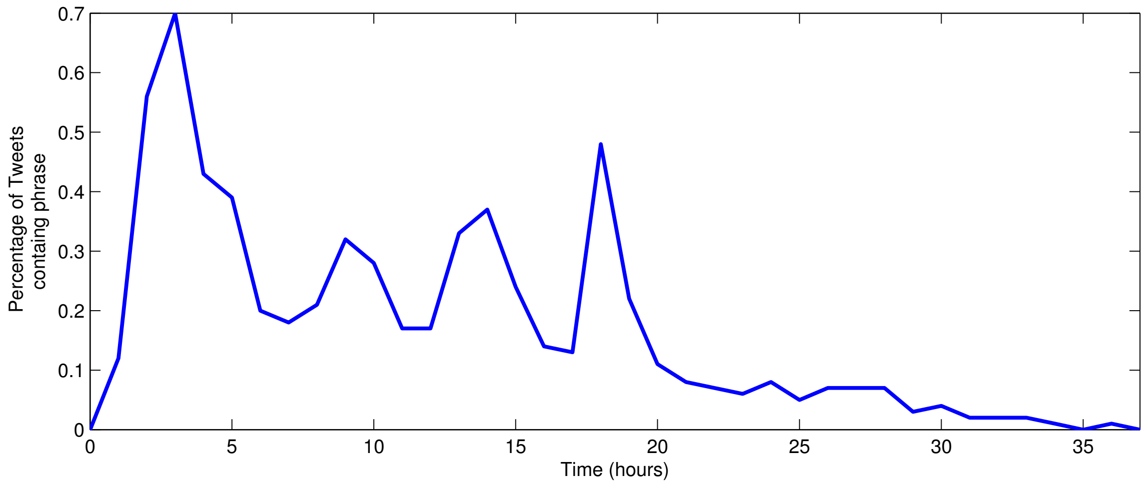

FIG. 9: Data for trend #WeAllNeed. This trend appears to display temporary oscillatory behaviour.

'preferential attachment' algorithm [13] where each new node is connected to m existing nodes with a probability proportional to the degree of the existing nodes. This generates a network with a power law distribution, $\sigma = 3$ and average connectivity of $2m$.

Scale-free networks are characterised by the relatively high prevalence of nodes of degree much greater than the average and a small average distance between any two nodes. This latter attribute is known as the small-world phenomenon which often appears in human social networks. The archetypal example of this is the six degrees of separation theorised to be between almost any two people on Earth.

Consider the Twitter network, when users join they are more likely to follow people with lots of followers and unlikely to follow people with very few. Popularity breeds popularity. This creates a network with the majority of users having only a few followers and a small minority of users having a very large number of followers. These highly influential users, called 'hubs' in network theory and 'super-spreaders' in epidemiology, should intuitively have a disproportionate effect on the success of a trend spreading, so they must be taken into consideration. As of 20th March 2012, Lady Gaga has the highest number of followers on Twitter with 21,041,668 [14], whereas the average user has only 126 followers [15].

The spreading of diseases on scale-free networks has been a focus of research due to its aptness for modelling sexually transmitted infections on the complex network of sexual partnerships. This is due to the extreme heterogeneity in the number of sexual partners people have. The majority of people have only a few sexual partners but there is a minority of people who are very sexually active and have a very large number of partners. This core group of high-risk individuals help maintain endemic infections in a population where the majority are in long-term monogamous relationships [16].

The internet and the world wide web have been shown to also exhibit scale-free characteristics [17]. The spread of computer viruses on the internet has been studied with interesting conclusions. Pastor-Satorras and Vespignani [17] used a susceptible-infected-susceptible model (where infected individuals recover to become susceptible again) and have shown that in this model epidemics do not have a threshold on an ideal (infinite population) scale-free network. In other words computer viruses can theoretically spread and become endemic even if they have an arbitrarily low transmission rate.

Since there are many parallels between these networks and Twitter's, the effect introducing this network structure may have on the SIR model should be considered.

This can be done by denoting the densities of susceptible, infected and recovered individuals of degree k at time t by $S_k(t)$, $I_k(t)$ and $R_k(t)$ respectively. We normalise by setting

$$S_k(t) + I_k(t) + R_k(t) = 1$$

It will be assumed that the network is uncorrelated. This means the probability a given neighbour of a node of degree i is of degree j depends only on the node-connectivity distribution. This probability is given by:

$$\frac{jP(j)}{\sum_k kP(k)}$$

where the sum is over all the values of node degree found in the network. The basic SIR system then becomes:

$$\frac{dS_k}{dt} = -S_k \sum_j \beta_{ij} I_j \quad (17)$$

$$\frac{dI_k}{dt} = S_k \sum_j \beta_{ij} I_j - \gamma I_k \quad (18)$$

$$\frac{dR_k}{dt} = \gamma I_k \quad (19)$$

where $\beta_{ij} = ij\beta/\langle k \rangle$ is the infection rate a susceptible node of degree i becomes infected by a node of degree j. $\langle k \rangle$ is the average degree of the nodes in the network calculated by:

$$\langle k \rangle = \sum_j jP(j)$$

once again summing over all values of node degree found in the network. It is possible to modify the models that have been discussed to include the network structure and thus create a more realistic but also more complicated models. Adding network structure made these models too complicated for analysis.

We define $\rho_0 = \beta D \langle k \rangle$ to be the average number of secondary cases an infected individual would create in an entirely susceptible homogeneous population, where D is some constant. By homogeneous we mean a network where every node is connected to $\langle k \rangle$ neighbours. The basic reproductive ratio now becomes:

$$R_0 = \rho_0 \frac{\langle k^2 \rangle}{\langle k \rangle^2}$$

For an ideal scale free network, that is to say one with infinite population, the variance of the connectivity distribution is infinite. This means

$$\frac{\langle k^2 \rangle}{\langle k \rangle^2} = \infty$$

as well which implies, regardless of the size of the transmission rate, $R_0 = \infty$. Hence the system has no epidemic threshold. For real world networks this will not be the case, however R_0 will still be much larger and the threshold will be lower than for the corresponding homogeneous network.

For our Twitter analogy this means a phrase theoretically has the potential to trend regardless of how 'infectious' it is. Let us suppose a hugely influential user becomes infected or starts an infection with a low transmission rate. Even though each of their users individually has a very low probability of becoming infected, because there are so many of them the phrase could still potentially trend. However we must factor in that the 'infectiousness' of a phrase may itself be increased just by being posted by a very popular user. This has been observed occurring with alarmingly high success for a member of the boy band One Direction, who regularly asks his followers to get a random hashtag of his choosing to trend.

R_∞ can be calculated by integrating Eq. (17) and Eq. (18) to get:

$$R_\infty = \langle 1 - e^{-k\alpha} \rangle$$

where

$$\alpha = \rho_0 \frac{\langle k(1 - e^{-k\alpha}) \rangle}{\langle k \rangle^2}$$

Compared to the final epidemic size for a homogeneous network R_∞ is lower [18]. This is a significant consequence for those wanting to know how many people will participate in a trend on Twitter.

Shaofen *et al.* introduced infection delay to the complex network SIR model [19]. This addition gives the following:

$$\frac{dS_k}{dt} = -S_k(t) \sum_j \beta_{ij} I_j(t - \tau) \quad (20)$$

$$\frac{dI_k}{dt} = S_k(t) \sum_j \beta_{ij} I_j(t - \tau) - \gamma I_k(t) \quad (21)$$

$$\frac{dR_k}{dt} = \gamma I_k(t) \quad (22)$$

where τ is some constant. They showed that when the transmission rate, β, is above a threshold a large delay naturally produces multiple epidemic peaks [19]. This may explain the multiple peaks that were observed in the trends in the data.

Another potential network expansion to the model is to add different transmission rates depending on the relationship between a follower and a followee. Huberman *et al.* have shown that there are two different networks at play on Twitter [20]. There is the large dense declared network of followers and the hidden sparser network of genuine social relationships. In other words most users follow many people but only actually directly interact on Twitter with a few. On Twitter it is possible to aim a post at a specific user by writing '@their_username' in a post. Around 25.4% of all posts are directed [20] so this is an important part of the Twitter experience.

This could be modelled on the network by adding different weightings of transmission rate to connections depending on the amount of direct interaction between two users. This can be done either by having a direct relation between transmission rate and level of interaction or more simply having two classes of connection, standard follower and friend. Huberman *et al.* defined a friend as someone a user has directed at least two posts to. This weak definition was enough to show a relationship between the amount of friends and level of Twitter usage for a user so this measurement is significant. This connection weighting would have to be directly calculated for a given network.

It is expected that a trend will be more likely to spread between friends compared to standard followers. This has significant consequences for those wanting to harness Twitter in order to spread an idea virally.

IX. DISCUSSION & CONCLUSION

Equation systems that can model the spread of topics on Twitter have been proposed taking into consideration relevant mechanisms. Trends have been observed which display popularity profiles that match those predicted by some of these models. Large statistical analysis is now needed to determine whether these equation systems are representative and actually due to the underlying mechanisms that have been modelled. This analysis is imperative in order to develop this research further.

The model can always be made to fit the data more closely by adding more variables and functions. This will be at the sacrifice of simplicity and make it more difficult to find analytical results. I have endeavoured at each stage to keep modifications simple and model only one aspect of Twitter functionality at a time.

The next logical step to develop this project further would be to make an agent based model. This would require writing a computer model that would generate Twitter-like networks of users who interacted following some set rules. It would then be possible to observe how a trend might propagate across this generated network and produce a much more realistic model. However, it would be hard to perform mathematical analysis on this model due to its complexity, although numerical analysis could prove to be fruitful.

There are some other aspects of modelling the Twitter network that could be considered in order to develop the model further. One of these would be to introduce stochasticity to the model. The models proposed in this project are deterministic mean field equations, which should be accurate for large populations such as the Twitter network. However if stochasticity has a powerful effect on the spread of trends the discussed models may not be accurate. Introducing stochasticity would require large amounts of statistics on the network but is an important element that needs to be considered.

It appears there are a great many factors that determine the success and behaviour of a trend. I believe that although these could potentially be quantified for a trend *post hoc* it would be very difficult to predict how a given phrase will perform, although some generalities may exist.

It has been discovered that there is a good deal of overlap between modelling the spread of diseases and modelling the spread of phrases on digital social networks. The techniques used extensively for disease models have potential for application in research into information dissemination on digital social networks, an area of increasing importance.

There are mechanisms exclusive to Twitter that have a significant effect on the behaviour of trend propagation. The mechanisms that have been determined to be most influential are the infection forcing from the Twitter trends table and extended activity by a minority of 'infected' users. The latter can explain the long tail often exhibited by Twitter trends. It has been discovered that the heterogeneous network structure of Twitter will have considerable effect on the spread of trends. A phrase does not need to be particularly 'infectious' to trend and the number of users who tweet on a trend will be lower than expected compared to an equivalent homogeneous network. It is also possible that adding delay to infection may model the multiple peaks observed in trends. Further research is required to establish whether this is the case.

The modelling of Twitter trends can reveal a great deal about the network behaviour of online social networks. It can also explain some observed phenomena. It may be possible to discern factors that do significantly effect the success of a trend or predict the nature of its spread. This will be of particular interest to groups who want to spread ideas, disseminate information and of course advertise virally on Twitter and similar networks.

X. APPENDIX

A. Glossary

API: Application Programming Interface. Source of all Twitter data and is used to build applications that access Twitter. Used by programs to access Twitter and has limitations for users that have not been 'white-listed' *i.e.* specifically given access by Twitter.

Follow: To follow someone on Twitter is to subscribe to their Tweets or updates on the site.

Follow count: The numbers that reflect how many people you follow, and how many people follow you.

Follower: A follower is a Twitter user who has subscribed to another user's Tweets.

Hashtag: The # symbol is used to mark keywords or topics in a Tweet e.g. #maths. It was created organically by Twitter users. Clicking on a hashtag will perform a search function for all Tweets containing the hashtag. Popular hashtags often become trending topics.

Meme: An element of a culture or behaviour that is passed from one individual to another by non-genetic means, especially imitation. On the internet it means an image, video or phrase *etc.* that is passed electronically from one internet user to another. Often these are internet in-jokes.

Retweet (noun): A Tweet by another user, forwarded to you by someone you follow. Often used to spread news or share valuable findings on Twitter.

Retweet (verb): To retweet, retweeting, retweeted. The act of forwarding another user's Tweet to all of your followers.

Timeline: A real-time list of Tweets on Twitter.

Trending topic: A subject algorithmically determined to be one of the most popular on Twitter at the moment.

Tweet (verb): Tweet, tweeting, tweeted. The act of posting a message, often called a "Tweet", on Twitter.

Tweet (noun): A message posted via Twitter containing 140 characters or fewer.

Tweeter: An account holder on Twitter who posts and reads Tweets. Also known as Twitterers.

Twitter: An information network made up of 140-character messages from all over the world.

B. Trend Data

Trend	Date Trended	Trendistic url
#sotellmewhy	02/01/12	http://trendistic.indextank.com/sotellmewhy/ _since-2012-01-02-10h-utc/_until-2012-01-03-10h-utc
#ISecretlyLove	04/01/12	http://trendistic.indextank.com/ISecretlyLove/ _since-2012-01-04-01h-utc/_until-2012-01-05-01h-utc
#inanafricanhouse	06/01/12	http://trendistic.indextank.com/inanafricanhouse/ _since-2012-01-05-16h-utc/_until-2012-01-07-05h-utc
#1ThingIFindSexy	09/01/12	http://trendistic.indextank.com/1ThingIFindSexy/ _since-2012-01-09-01h-utc/_until-2012-01-10-11h-utc
sloot	11/01/12	http://trendistic.indextank.com/Sloot/ _since-2012-01-11-03h-utc/_until-2012-01-15-02h-utc
#WeAllNeed	11/01/12	http://trendistic.indextank.com/WeAllNeed/ _since-2012-01-11-01h-utc/_until-2012-01-12-22h-utc
#youknowshestheone	25/02/12	http://trendistic.indextank.com/youknowshestheone/ _since-2012-02-25-01h-utc/_until-2012-02-26-01h-utc
#somewhereintheghetto	27/02/12	http://trendistic.indextank.com/somewhereintheghetto/ _since-2012-02-27-01h-utc/_until-2012-02-28-11h-utc
#Iwannagiveashoutoutto	27/02/12	http://trendistic.indextank.com/iwannagiveashoutoutto/ _since-2012-02-27-23h-utc/_until-2012-02-29-23h-utc
liam payne	15/03/12	http://trendistic.indextank.com/liam-payne/ _since-2012-03-15-00h-utc/_until-2012-03-15-13h-utc

ACKNOWLEDGEMENTS

I would like to thank my supervisor Colm Connaughton for his guidance and contribution. This project would not have been possible without him. I would also like to thank Matthew Keeling who taught the module Population Dynamics. His notes proved invaluable for the disease models I have used. Thanks must also be given to Twitter for such a truly fascinating website that this project was only able to scrape the surface of.

[1] L. Grossman, Time Magazine (2009).
[2] BBC, "England riots: Twitter and Facebook users plan clean-up," http://www.bbc.co.uk/news/uk-england-london-14456857 (2011).
[3] Daily Mail, "Gagging law stars 'outed' on Twitter: Thousands see the names of celebrities alleged to have taken out injunctions," http://www.dailymail.co.uk/news/article-1384883/Super-injunctions-Twitter-user-outs-gagging-order-celebrities-thousands.html (2011).
[4] Trendistic, http://trendistic.indextank.com/ (2012).
[5] Janette Lehmann, Bruno Gonçalves, José J. Ramasco, and Ciro Cattuto, WWW'12 (2012).
[6] Sitaram Asur, Bernardo A. Huberman, Gabor Szabo, and Chunyan Wang, http://dx.doi.org/10.2139/ssrn.1755748 (2011).
[7] Romualdo Pastor-Satorras and Alessandro Vespignani, Phys. Rev. E **63** (2001).
[8] W. O. Kermack and A. G. McKendrick, Proc. Roy. Soc. Lond. **A**, 700 (1927).
[9] D. J. Daley and J. Gani, Cambridge University Press (2005).
[10] Ottar N. Bjørnstad, Barbel F. Finkenstadt, and Bryan T. Grenfell, Ecological Monographs **72**, 169 (2002).
[11] Buzzgain, "How many tweets does it take to be a trending topic on Twitter?" http://news.buzzgain.com/how-many-tweets-does-it-take-to-be-a-trending-topic-on-twitter/ (2009).
[12] Tao Zhou, Matus Medo, Giulio Cimini, Zi-Ke Zhang, and Yi-Cheng Zhang, PLoS ONE **6**, e20648 (2011).
[13] Albert-László Barabási and Réka Albert, Science **286**, 509 (1999).
[14] Twitaholic, http://twitaholic.com/ (2012).
[15] Evan Weaver, "Twitter, an Evolving Architecture," http://www.infoq.com/news/2009/06/Twitter-Architecture (2009).
[16] Matt J Keeling and Ken T.D Eames, J. R. Soc. Interface **2**, 295 (2005).
[17] Romualdo Pastor-Satorras and Alessandro Vespignani, Phys. Rev. Lett. **86** (2001).
[18] Robert M. May and Alun L. Lloyd, Phys. Rev. E **64** (2001).
[19] Shaofen Zou, Jianhong Wu, and Yuming Chen, Phys. Rev. E **83** (2011).
[20] Bernardo A. Huberman, Daniel M. Romero, and Fang Wu, First Monday [Online] **14** (2008).