

Elements of computer modelling: 12 studies for beginners.

Study 3: Rumour propagation model

Illia O. Teplytskyi

Kyryvi Rih State Pedagogical University, 54 Universytetskyi Ave., Kyryvi Rih, 50086, Ukraine

Abstract. This paper continues our foundational series on computer modelling [1, 2] and introduces the first complete modelling exercise: a rumour propagation model through a community. Beginning with a deliberately oversimplified version that produces unlimited geometric growth, we progressively refine the model across four versions: incorporating varying communication levels among population members (ordinary speakers, talkative individuals, and silent members), accounting for repeat knowers who have already heard the rumour, and finally developing an analogous epidemic model using differential equations. Each version follows the complete modelling cycle established in the preceding studies: problem formulation, simplifying assumptions, mathematical model construction, spreadsheet algorithm development, computational experiment, and critical analysis of results. The progressive refinement demonstrates how identifying a model's weaknesses – particularly overly simplistic assumptions – drives the creation of more realistic versions. Key concepts introduced include model stability, the distinction between artificial and natural growth limitations, and the importance of model testing. The epidemic model in the final section provides an independent exercise that reinforces all acquired skills whilst introducing continuous-time modelling with finite differences.

Keywords: rumour propagation, discrete models, model refinement, geometric progression, computational experiment, spreadsheet modelling, model stability, epidemic model, finite differences, communication modelling

Section II. First models

We shall begin the direct study of computer modelling by examining a problem that does not require specialised knowledge from any particular subject domain. Using this example, we shall obtain initial concepts about the main stages and some common techniques involved in creating and investigating computer models.

The proposed example is the problem of rumour propagation. Admittedly, few people are troubled by problems connected with rumour spreading; however, like most problems that genuinely arise in practice, it is an example of a typical modelling problem – namely, *a problem with an imprecisely formulated condition*, which substantially distinguishes it from the usual problems found in school textbooks.

To consolidate the skills acquired whilst working on the rumour problem, immediately following it we propose the problem of creating and investigating a simple epidemic model. Unlike the commentaries on the rumour problem, here only textual guidance is provided and sample tables and graphs are entirely absent. You will have to perform this part of the work independently, using the solutions from the previous problem as templates. At the end, you will need to think through and implement more sophisticated versions of this model.

Let us note at once that, in terms of the ideas proposed, the approaches to problem solving considered below are not the only possible ones, and this multiplicity of approaches is also characteristic of problems in computer modelling.

 0000-0002-9416-3630 (I. O. Teplytskyi)

 ilia2306@gmail.com (I. O. Teplytskyi)

 <https://scholar.google.com/citations?huser=f2BIZ7EAAAAJ> (I. O. Teplytskyi)

Received	Accepted	Published	Version of record
2026-01-06	2026-06-23	2026-07-25	2026-07-25



Science
Education
Quarterly



© Copyright for this article by its authors, published by the Academy of Cognitive and Natural Sciences. This is an Open Access article distributed under the terms of the Creative Commons License Attribution 4.0 International (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

3. Rumour propagation model

Problem formulation

Modern mass media (radio, television, the Internet) make it possible to deliver any news anywhere in a short time. At the same time, the ancient method of transmitting information – from person to person – still exists today, and it mostly concerns rumours. For all its apparent imperfection, this method can today be so effective (mobile communication, email, video calls) that it is capable of competing even with television.

Let us try to model the process of rumour propagation within a given community. As the goal of modelling, let us set the task of finding the answer to the question: *how many community members will know the news after a given period of time, if at the initial moment it is known to only a few of them?*

Let us outline the general plan for further work.

1. To formalise the problem conditions and create its conceptual model, we shall introduce quantitative characteristics (parameters) of the process. Establishing relationships between these characteristics in the form of mathematical expressions will allow us to create a mathematical model.
2. Initially, we shall consider the simplest version of the model, which we shall subsequently improve. We shall conclude each successive version with an analysis of the results obtained. The purpose of such analysis is to identify the weakest points in the model created and to search for ways to overcome them. Usually, the source of weak points turns out to be overly simplified assumptions, and it is precisely these that will need to be replaced with more complex ones, in the hope that the modelling results will prove more realistic and adequate.
3. For each version, we shall compose a working algorithm and, in accordance with the mathematical model, create a computer model.

Analysis of the problem: features of the process

For convenience, we shall use the words “rumour” and “news” as synonyms, and call the process of communicating them “transmission”.

1. For different narrators, the intervals between successive transmissions, as well as the duration of these transmissions, may take arbitrary random values.
2. Among listeners, there are those to whom the news is as yet unknown and those who already know it – the “*knowers*”. During transmission, they “receive” the news a second time. It is entirely natural that under such circumstances, some of these repeat knowers may either lose interest in further spreading the news or continue to spread it.
3. People differ in their ability to communicate (level of sociability), and the more talkative among them (let us call them *talkers*) under equal conditions communicate the news to a greater number of listeners than others – ordinary narrators. It is known that some people do not like to tell stories at all. Let us call them *silent ones*.
4. With each successive transmission, the number of knowers of the news increases, and given a constant numerical composition of community members, the search for new listeners over time turns into a genuine problem – the problem of hunting for “free ears”. Eventually, one of two things must happen: either the spreading of the rumour ceases, or the rumour “circulates” among the community members on the principle “I know that you know, but listen anyway ...”.

Each of the factors just listed is capable of substantially complicating the study of the process, and their simultaneous influence goes without saying.

3.1. Simplest version of the model

To simplify the situation, let us adopt a number of assumptions.

Assumption 1. Transmission sessions begin simultaneously for all narrators and have equal duration.

Assumption 2. Every knower of the rumour spreads it throughout the entire modelling time we have chosen, without introducing any distortions; otherwise it would already be a different rumour.

Assumption 3. At each successive transmission, the rumour should reach only those persons who have not previously heard it. To fulfil this condition, before transmitting it one should first obtain a negative answer to the question: “Have you heard that...?”

Remark. In reality, rumour spreading does not occur around the clock: the overwhelming majority of people have many other important things to do, and most often people communicate not for the purpose of listening to and spreading rumours. Let us recall that we are discussing the rumour problem as a simple example of an approach to creating and investigating a model.

It appears more convenient to count the number of knowers of the rumour not after a certain time, but after a certain number of transmission sessions. In that case, it is advisable to reformulate the modelling goal as follows:

How many community members will know the rumour after a certain number of transmission sessions, if at the initial moment it is known to only a few of them?

Let us formalise the problem conditions. This means that, being at the stage of problem formulation, we shall transform the conditions so that they preclude ambiguity, that is, they should have a unique interpretation. Let us introduce the following quantitative characteristics:

j – the sequential number of the transmission session ($j = 0, 1, 2, \dots, m$);

N_0 – the initial ($j = 0$) number of knowers of the rumour;

N_j – the number of knowers after the session numbered j .

Let each knower during one transmission session communicate the rumour to k new persons. In that case, to satisfy Assumption 3, it is necessary that during the transmission session each knower finds k persons to whom the rumour is still unknown.

Here N , j , k are positive integers.

As noted above, the parameter k for different people has different values depending on their individual level of sociability and, furthermore, under different circumstances k may randomly change over time. In connection with this, let us adopt one more

Assumption 4. We shall consider the parameter k to be constant over time and identical for all knowers.

Then, for example, $k = 2$ means that throughout the entire modelling time each knower during one transmission session tells the rumour to two new persons, and each of these two in the next session passes the rumour to two others, and so on. In this case, with $N_0 = 1$:

- after the first session, the number of knowers becomes: 1 (initial) + 2 (new) = 3;
- after the second session, each of these three tells the rumour to two new persons, and the number of knowers increases by $3 \times 2 = 6$ during the second session, giving a total of $3 + 6 = 9$ after two sessions, and so on.

If after the session numbered $(j - 1)$ the rumour is known to N_{j-1} persons, then after the next session numbered j the number of knowers will increase by

$$\Delta N = N_{j-1} \cdot k, \quad (1)$$

and their total number will be

$$N_j = N_{j-1} + \Delta N, \quad (2)$$

which, taking (1) into account, gives

$$N_j = N_{j-1}(1 + k). \quad (3)$$

The systems of equations (1) and (2), or (3), constitute the simplest mathematical model of our rumour propagation problem.

Remark. In this model, by agreement, time t does not appear explicitly, that is, equations (1) and (2) do not contain the variable t , but if necessary, the time can always be found: $t = j \cdot \Delta t$, where Δt is the duration of one transmission session.

Thus, the problem conditions have been formalised and its mathematical model has been created.

Computer model. Algorithm for working with the model

The computer model of this problem is formed by a spreadsheet whose cells contain formulae in accordance with the mathematical model. The working algorithm contains the following instructions.

1. Create a spreadsheet following the template and fill its first row with variable names, and in column D (“Given:”) enter the names of model parameters N_0 and k together with the “=” signs.

	A	B	C	D	E
1	j	ΔN	N	Given:	
2				$N_0 =$	
3				$k =$	
...	...	...	...		

2. Enter numerical values of input data into cells E2 and E3.
For example, 1 and 2, respectively ($N_0 = 1$ and $k = 2$).
3. Fill the second row with initial values of variables:
 - 3.1. $j = 0$ – the initial session number;
 - 3.2. $\Delta N = 0$ – before the first session, the increment ΔN of the number of knowers equals zero;
 - 3.3. $N = N_0$ – by reference to cell E2 from “Given:”.

The specified cells of the table will have the following content:

Cell	Formula/Number	Comments
A2	0	
B2	0	
C2	=E2	reference to N_0

4. Fill the third row, that is, write formulae corresponding to equations (1) and (2) for $j = j + 1$ – the number of the next transmission session:

Cell	Formula	Comments
A3	=A2+1	next session number
B3	=C2*\$E\$3	ΔN by equation (1)
C3	=C2+B3	N by equation (2)

5. Repeat the commands of the third row n times, where n is the number of transmissions being modelled (in our example, $n = 20$). To do this, copy the formulae of the third row into the subsequent $n = 20$ rows.
6. Using the table data, plot graphs of the dependencies $\Delta N = \Delta N(j)$ and $N = N(j)$.

Computational experiment

Suppose, for example, that in a community where all inhabitants have a transmission parameter $k = 2$, $N_0 = 1$ knower of a piece of news appears, having the same transmission parameter. How many new people will learn about the news after $n = 20$ transmission sessions?

Substituting the input data into cells E2 and E3, we obtain the table, a fragment of which is shown in figure 1. Here we see that as the number of transmissions increases, the number of knowers grows steeply. From the table it is also clear that successive values of the increment ΔN and the number of knowers N form geometric progressions with common ratio 3.

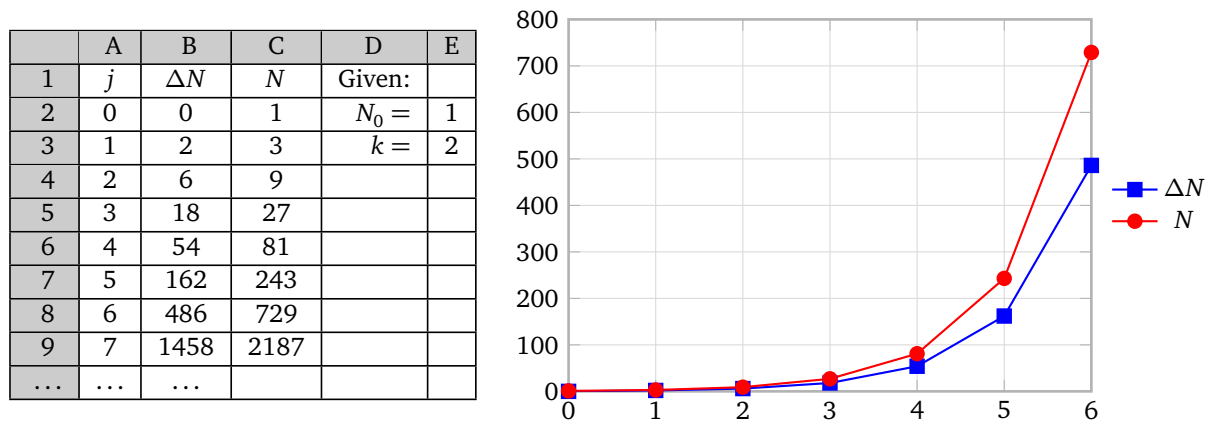


Figure 1: Fragment of the spreadsheet and related plot for the simplest rumour model ($k = 2$, $N_0 = 1$).

Remark. The variables N and ΔN are by their meaning integers. One should bear in mind that spreadsheets allow real numbers to be visually displayed in integer format, taking rounding rules into account. However, in the computer's memory, these numbers will remain in real number format with the maximum possible number of digits for the given spreadsheet processor. In all arithmetic operations, they will participate precisely in real number format.

Exercise

1. To verify the last *Remark*, enter the expression `=22/7` into any free cell.
 - 1.1. Display the result in “Integer” format (a number with zero decimal places).
 - 1.2. Next, copy the contents of this cell without formulae into some other free cell and increase the size of this cell so as to see all decimal places.
2. How many rows of the table in figure 1 were used to plot the corresponding graph? Which rows are these?
3. Analysing expressions (1), (2), and (3), answer the question: why, with a transmission parameter value $k = 2$, do both progressions of ΔN and N values have a common ratio of 3?
4. With $N_0 = 1$, after the sixth transmission ($j = 6$), the number of knowers is 729. If the initial number of knowers is increased tenfold (that is, take $N_0 = 10$), can one assert that the number of transmissions j required for the same number of people (729) to learn the rumour will also decrease tenfold compared to the case $N_0 = 1$?

In this version of the model, the news spreading process occurs so rapidly that after just the 13th transmission the number of knowers exceeds 2 million, and after two or three more transmissions it becomes greater than the population of any city in the world: the geometric progression demonstrates its steep character.

If we take the number of transmissions per day to be 10, the noted effect will be achieved in less than two days. Clearly, such a course of events does not correspond to reality, meaning the constructed model is not adequate to the real situation and, consequently, there arises a need to improve the model. But before proceeding to create a more sophisticated model, let us present some considerations.

Let us enquire how many transmission sessions must occur for the rumour to reach all members of the community. Here, a natural desire immediately arises to introduce the total number of community members into consideration. Let us adopt another

Assumption 5. Let the numerical composition of the given group remain constant throughout the modelling time (among the community members there are no deaths, no isolation, no one leaves and no one arrives). For not too lengthy time intervals and small groups, this is acceptable.

Now let us make the information displayed on the screen more convenient for viewing and analysis. To do this, we shall artificially limit the number of knowers N_j so that it does not exceed the total population S .

For this purpose, let us make some changes to the algorithm:

— item 2: add to the input data a new variable S (cell D4) – for the population size (e.g., 10 000 – cell E4).

Cell	Formula/Number
D4	S
E4	10000

— item 4.2: compute ΔN according to expression (1);

if the number of knowers N_j is not less than S **then**

in the *next* row perform the assignment

$$\Delta N = S - N_{j-1}$$

else

leave ΔN as before

end if

— item 4.3:

if $N_{j-1} + \Delta N$ is not less than S **then**

in the next row perform the assignment

$$N_j = S$$

else

$$N_j = N_{j-1} + \Delta N$$

end if

In the table, this is implemented by the following formula:

Cell	Formula
B3	=IF(C2*\$E\$3>=\$E\$4,\$E\$4-C2,C2*\$E\$3)
C3	=IF(B3=0,\$E\$4,C2+B3)

Do not forget(!) to copy the formulae from B3 and C3 to other cells in their columns.

Reviewing the new table (figure 2) shows that the stated goal has been achieved: now a single glance suffices to establish immediately after which transmission session the entire population becomes informed.

Comment:

- 1) the content of highlighted cells in the figure 2 (B10, B11, B12);
- 2) the curve for ΔN , using data from this table from columns B and C, as well as the graph $N = N(j)$;
- 3) the “bumps” on the graph for ΔN at $j = 10$ and on the graph for N at $j = 9$ (figure 4a). How can one eliminate these “bumps” and obtain figure 4b?

Conclusions

1. The version of the model considered is very simplified. The Assumptions 1–5 underlying it lead to a result that could have been predicted in advance even without modelling. Namely: the number of knowers, as the number of transmissions increases, grows rapidly and without bound.

	A	B	C	D	E
1	j	ΔN	N	Given:	
2	0	0	1	$N_0 =$	1
3	1	2	3	$k =$	2
4	2	6	9	$S =$	10000
5	3	18	27		
6	4	54	81		
7	5	162	243		
8	6	486	729		
9	7	1458	2187		
10	8	4374	6561		
11	9	3439	10000		
12	10	0	10000		
...	...	...			

Figure 2: Fragment of the spreadsheet for the simplest rumour model ($S=10000$, $k=2$, $N_0=1$).

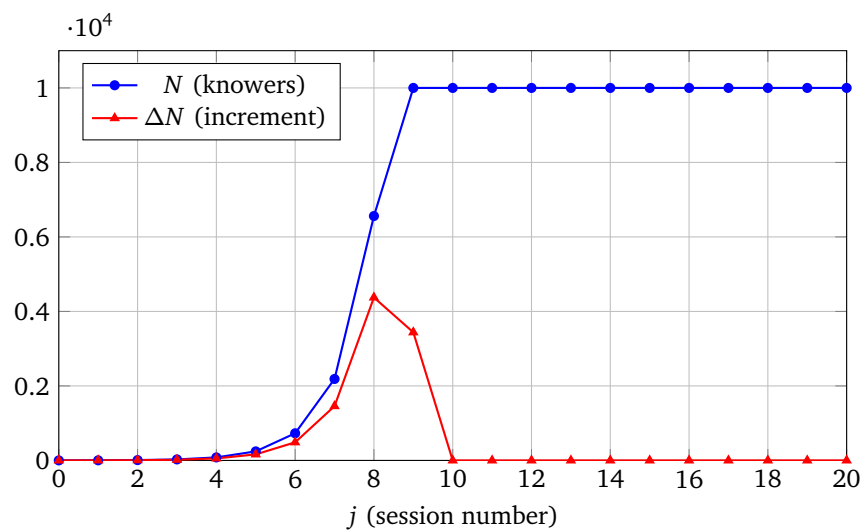


Figure 3: Graphs of $N = N(j)$ and $\Delta N = \Delta N(j)$ for the simplest rumour model with population limit $S = 10000$, $k = 2$, $N_0 = 1$.

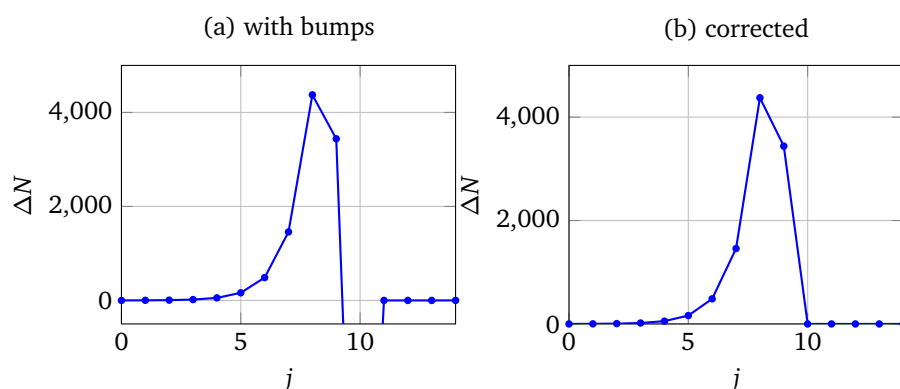


Figure 4: The $\Delta N = \Delta N(j)$ graph: (a) with “bumps” due to the limiting condition; (b) corrected version.

2. The introduction of a limitation on the number of knowers does not fundamentally make the model more reliable. It is clear that the measures taken for this purpose are of a frankly artificial character: the limitation of the number of knowers in no way follows from the model itself (from the system of equations).

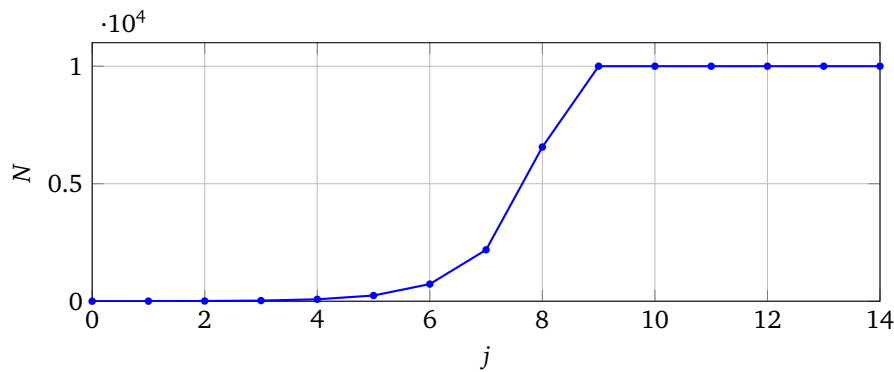


Figure 5: Corrected $N = N(j)$ graph without the bump artefact at $j = 9$.

3. At this stage of the work, we should not seriously discuss the question of correspondence between the model and the given process, that is, it is not worth discussing the question of model adequacy at this point: if it exists at all, it is extremely weak.

Remark. When discussing the question of mathematical model adequacy, one should bear in mind not adequacy in general (no such thing exists), but adequacy with respect to certain features or properties of the object.

4. The principal value of this version lies in the fact that it can serve as a foundation for further improvement of the model.
5. According to item 1 of the computational experiment, expression (3)

$$N_j = N_{j-1}(1 + k)$$

defines a geometric progression. The sum of its first j elements can be found using a well-known formula. *Which one?*

There exist quite a few objects described by such a formula exactly, yet they have nothing to do with rumour spreading. In particular, if N_0 is the size of an initial monetary deposit and k is the annual bank interest rate, then $S_j = N_0 \cdot (1 + k)^j$ is the size of the deposit after j years. In this example, j is necessarily a natural number. In a number of other cases, j may take real values.

Try to give your own similar example.

3.2. Improved model version: accounting for the communication level of community members

The assumptions used in our previous version were too simple, so the results of working with the model turned out to be practically uninteresting. Let us proceed to improving the model.

According to Assumption 4, our model assumes a population uniform with respect to talkativeness. In real life, this usually does not occur. Therefore, first and foremost, let us remove this restriction and formulate

Assumption 4*. Let the population consist of three groups:

- 1) ordinary people with transmission parameter k_{ord} ;
- 2) talkers with transmission parameter k_t ($k_t > k_{\text{ord}}$);
- 3) silent ones with transmission parameter $k_s = 0$.

As before, the values of these parameters are considered constant over time.

Let the relative share (percentage) of silent ones be M , and of talkers B , of the total number of community members. Then the relative share Z of ordinary knowers will be

$$Z = 1 - M - B.$$

It should not provoke serious objection to assert that with a sufficiently large population size, the corresponding shares (percentages) are distributed in the same way among the knowers of the news. Thus, for example, if the percentage of silent ones among the entire population is M , then among the knowers their percentage is also M .

Now the increment ΔN of the number of knowers after the transmission session numbered j will be provided by the talkers and ordinary narrators:

$$\Delta N = \underbrace{N_{j-1} \cdot B \cdot k_t}_{\text{due to talkers}} + \underbrace{N_{j-1} \cdot k_{\text{ord}} \cdot (1 - M - B)}_{\text{due to ordinary}} \quad (4)$$

or

$$\Delta N = N_{j-1} \cdot (B \cdot k_t + k_{\text{ord}} \cdot Z); \quad (5)$$

the total number of knowers will be sought according to (2):

$$N_j = N_{j-1} + \Delta N,$$

or, taking (5) into account,

$$N_j = N_{j-1}(1 + B \cdot k_t + k_{\text{ord}} \cdot Z). \quad (5^*)$$

Equations (2) and (5), or (2) and (5*), are the improved mathematical model of the process.

The modified table has the following layout:

	A	B	C	D	E
1	j	ΔN	N	Given:	
2				$M =$	
3				$B =$	
4				$k_t =$	
5				$k_{\text{ord}} =$	
6				$N_0 =$	
...	...	...	...		

In accordance with the new equations, certain changes are also required in the algorithm.

Algorithm modifications

— item 2: to the input data, instead of the former k , add the parameters M , B , k_t , and k_{ord} (cells D2:D5 and E2:E5, respectively), and move the parameter N_0 to cells D6 and E6.

— item 4.2: compute the increment ΔN of the number of knowers according to (5*):

Cell	Formula
B3	$=C2*($E$3*$E$4+$E$5*(1-$E$2-$E$3))$

The values of parameters M and B must satisfy the condition $M + B \leq 1$. To at least partially prevent incorrect entry of these parameter values, after filling cells E2 and E3 we shall display a corresponding message on screen. For this purpose, the following formula should be created in cell E8:

$$=IF(E2+E3>1, "M+B must be <=1!", "").$$

Other items of the algorithm remain unchanged.

Computational experiment

1. For definiteness, let us assume that by the sociability trait, the community members are distributed as follows: the relative share of silent ones M is 0.75 (75%), the relative share of talkers B is 0.1 (10%), the relative share of ordinary people $Z = 1 - M - B$. The transmission parameters $k_t = 3$, $k_{ord} = 2$.

Among the community members, $N_0 = 1$ knower of the rumour appears. How many people will learn about it after $n = 22$ transmission sessions?

2. Enter the data and obtain the table (figure 6).

	A	B	C	D	E
1	j	ΔN	N	Given:	
2	0	0	1	$M =$	0.75
3	1	1	2	$B =$	0.10
4	2	1	3	$k_t =$	3
5	3	2	4	$k_{ord} =$	2
6	4	2	7	$N_0 =$	1
7	5	4	10		
8	6	6	17		
9	7	10	27		
...	...	...	...		

Figure 6: Fragment of the improved model spreadsheet ($M = 0.75$, $B = 0.1$, $k_t = 3$, $k_{ord} = 2$, $N_0 = 1$).

- 2.1. Does the sequence of values of the variable N_j form a geometric progression? Why?

If your answer is negative, then you are mistaken. You may have been misled by the formulation of this question regarding a sequence of such an unusual kind. First, one should not forget that the numbers in the table cells are represented in integer format only visually, whereas in the computer's memory they are stored in real number format. Second, look at expression (5*): from it we see that each subsequent element of the sequence N_j equals the previous one N_{j-1} multiplied by the constant number $q = 1 + k_t \cdot B + k_{ord} \cdot Z$. Therefore, the sequence N_j is a geometric progression, and the number q is its common ratio.

- 2.2. Write the formula for an arbitrary element of the formed sequence in the form $a_n = a_1 \cdot q^{n-1}$. For this purpose, using our data, find the specific values of N_1 (in place of a_1), q , and n .
- 2.3. What form does the expression for N_j now take?
- 2.4. Create a new column in the table after the column for N values, in which the values of N_j will be computed according to the expression from 2.2: $N_j = N_1 \cdot (1 + k_t \cdot B + k_{ord} \cdot Z)^{j-1}$. For this purpose, enter the following expressions into cells D2 and D3:

Cell	Formula	Comments
D2	=F\$6	N_0
D3	=C\$3*(1+F\$3*F\$4+F\$5*(1-F\$2-F\$3))^(A3-1)	N_j by formula from 2.2

- 2.5. Compare the values of N_j in columns C and D and draw conclusions.
- 2.6. One can also verify that the sequence N_j is a geometric progression in another way – proceeding from the definition of such a sequence, namely: the ratios N_j/N_{j-1} must be constant.

Delete the previous values from column D and in cell D3 create the appropriate formula ($=C3/C2$), which should be copied to other cells in column D. How should one explain the fact that the values in all cells of column D turn out to be identical?

Now delete column D from the table.

3. Determine how the displayed values of N_j are affected by changing the parameters M and B . To this end, compare the results with the initial data against the results with:
 - 3.1. $M = 0.75$ and $B = 0.05$;
 - 3.2. $M = 0.8$ and $B = 0.1$.

Explain the reasons for the observed changes.

4. Restore the initial values of parameters M and B . Keeping them unchanged, try simultaneously to select for the transmission parameters k_{ord} and k_t such new values that the previous results remain in column C.

When you tire of this activity, enter $k_{\text{ord}} = 1$ and $k_t = 4.5$ ($k_t = 4.5$ can be interpreted as follows: each talker over two transmission sessions finds 9 listeners).

- 4.1. Where did these numbers 1 and 4.5 come from? Do other such pairs exist? How many of them can there be? In other words, does a relationship exist between the transmission parameters k_{ord} and k_t under the condition of item 4?
- 4.2. Reliable answers to all these questions can be given if we again turn to expression (5*), which was rewritten in the form

$$N_j = N_{j-1} \cdot q, \quad \text{where} \quad q = 1 + B \cdot k_t + k_{\text{ord}} \cdot Z.$$

- 4.3. It is clear that the same values of N_j will appear every time the same values of q are ensured: $q_1 = q_2$, that is,

$$B \cdot k_{t1} + k_{\text{ord}1} \cdot Z = B \cdot k_{t2} + k_{\text{ord}2} \cdot Z.$$

This equation simultaneously contains both unknowns k_{t2} and $k_{\text{ord}2}$. Solving it, for example, with respect to k_{t2} , we have

$$k_{t2} = k_{t1} + \frac{Z}{B}(k_{\text{ord}1} - k_{\text{ord}2}).$$

Taking further, according to the condition, $M = 0.75$, $B = 0.1$, $k_{\text{ord}1} = 2$, $k_{t1} = 3$, we obtain

$$k_{t2} = 6 - 1.5 \cdot k_{\text{ord}2}.$$

Now one can assign certain values to $k_{\text{ord}2}$ and obtain corresponding values for k_{t2} . It is only necessary to control oneself with the obvious restrictions $k_{t2} > 0$, $k_{\text{ord}2} > 0$, $k_{t2} > k_{\text{ord}2}$.

Exercise

1. Answer the questions from item 4.1.
2. As in item 3 of the computational experiment of the previous model version, create a limitation on the number of knowers N_j so that it again does not exceed the total population S of the community, and compare the new table with the table of the previous version.
3. Under what conditions can this version of the model be reduced to the previous one?

Conclusions

1. The improved version of the model is a logical continuation of the previous one and contains it as a special case. It is somewhat more interesting and realistic, but still remains very simplified and represents only an insignificant step forward.
2. The search for answers to the questions posed in item 4 during the computational experiment was aimed at demonstrating one of the important features of the researcher's work in the modelling process. Its essence lies in the fact that, whilst performing a computational experiment and analysing its results, one can, of course, try to search for answers by "blind" trial and error,

but attempts to solve the problem by analysing the mathematical relationships underlying the model are significantly more effective. However, as we shall see later, in modelling there frequently occur cases when the researcher is deprived of such a possibility, and then the only remaining path is trial and error.

3. Again, as in the previous version of the model, the introduction of a knower number limitation does not make the model more reliable. The reason is the same – the limitation on the number of knowers is *artificial*, since it does not at all follow from the model itself.
4. In both versions of the model considered, the weakest point is the requirement formulated in Assumption 3. Indeed, must the rumour necessarily reach only those people to whom it is unknown? If this requirement is removed, it will mean that listeners may also include community members to whom the rumour is already known (*repeat knowers*). Moreover, in real life this is precisely what often happens.

3.3. Further model refinement: accounting for repeat knowers

Let us now discuss the method of counting the number X of repeat knowers. Even if in the next session we prevent the return of information from the freshly created knower of the news to his or her immediate informant, the information can still return to this same informant from one of the other knowers. This means that excluding the mentioned possibility is not fundamental, that is, already in the very next transmission session the rumour may return to its source from the previous session.

Moreover, repeat knowers can appear during the very first transmission if $N_0 > 1$ (some of the initial knowers may become the listener of another initial knower). Thus, counting the number X of repeat knowers can begin from the very first transmission session.

Let us recall a familiar situation: during a break, a small group of pupils has gathered to listen to news from a classmate. Everyone is interested except one. This one is indifferent – the rumour being spread is already known to him. He, like any other knower, is himself always ready to pass it on to anyone at the opportunity: after all, according to Assumption 2, every(!) knower spreads information throughout the entire modelling time. How many *new* knowers appear as a result of the considered transmission? Clearly, one fewer than there were listeners in the group.

Thus, by cancelling Assumption 3, we find ourselves faced with the necessity of not counting as new knowers those persons for whom the rumour did not become new, that is, those who received it a second time. And not because in life mostly only new information is of interest, but precisely because repeat knowers do not increase the total number of knowers. To account for the number of such persons, various ideas can be used. Let us consider one of the possible approaches.

We shall assume that at the very beginning of the spreading process, the rumour is new for all S members of the community. After the first transmission, the *total* number of knowers (those for whom the rumour is new together with those to whom it is already known) will increase by Δn (for the increment of only new knowers we shall, as before, retain the notation ΔN). Let us note that repeat knowers can be found only within the increment Δn , which we shall compute using expression (3) or (5*) from the previous version.

The expression $\Delta n/S$ has the meaning of that fraction of the entire population within which it is advisable to search for repeat knowers, and $Q = X/\Delta n$ is the fraction of repeat knowers in the current increment ($Q \leq 1$).

We shall search for X in two steps.

Step 1. First, let us adopt

Assumption 6. The fraction Q of persons who during the current transmission receive the information a second time is proportional to $\Delta n/S$:

$$Q = G \cdot \frac{\Delta n}{S}, \quad (6)$$

where G is a coefficient whose value we shall select *experimentally*.

Step 2. Next, compute the absolute number X of persons who receive information a second time:

$$X = Q \cdot \Delta n. \quad (7)$$

Then the actual increment ΔN of the number of persons who receive the news for the first time will be

$$\Delta N = \Delta n - X. \quad (8)$$

Acting according to (6)–(8) at each successive transmission, we shall continue to find the increment ΔN of only those persons who receive the news for the first time. Computing further

$$N_j = N_{j-1} + \Delta N,$$

we shall be able to obtain the total number of precisely those knowers who will spread the news in the next transmission.

Algorithm modifications

— items 1, 2: move the input data from column E to column H; add to the input data the variables S and G (cells H2 and H8, respectively; place other parameters as before in H3:H7).

In three new columns D, E, and F, we shall observe the “behaviour” of the new variables Δn , Q , and X , respectively.

— item 4.2:

$Z = 1 - M - B$ – the relative share of ordinary narrators;

$\Delta n = N_{j-1} \cdot (B \cdot k_t + Z \cdot k_{ord})$ – as in the previous model version, but here Δn is the temporary value of the variable ΔN ;

$Q = G \cdot \Delta n / S$ – the relative share of persons who receive the rumour a second time;

G – a coefficient selected *experimentally*;

$X = Q \cdot \Delta n$ – the absolute number of persons who during the current transmission receive the rumour a second time;

$\Delta N = \Delta n - X$ – the final increment of the number of new knowers (only those to whom the rumour comes for the first time).

The corresponding table cells have the following content:

Cell	Formula	Comments
A3	=A2+1	j
D3	=C2*(\$H\$4*\$H\$5+\$H\$6*(1-\$H\$3-\$H\$4))	Δn
E3	=\$H\$8*D3/\$H\$2	Q
F3	=E3*D3	X
B3	=D3-F3	$\Delta N = \Delta n - X$

— item 4.3 remains unchanged: $N_j = N_{j-1} + \Delta N$; here N_j is the final number of knowers capable of spreading the rumour after session number j . In the table, as in the previous model version, this is implemented as follows:

Cell	Formula
C3	=C2+B3

— item 5 also remains unchanged.

— Copy the formulae of the third row (cells A3, B3, C3, D3, E3, and F3) into other cells of columns A, B, C, D, E, and F, respectively.

	A	B	C	D	E	F	G	H
1	j	ΔN	N	Δn	X	Q	Given:	
2	0	0	1	0	0	0.000000	$S =$	10000
3	1	1	2	1	0	0.000135	$M =$	0.6
4	2	2	4	2	0	0.000256	$B =$	0.05
5	3	3	7	3	0	0.000487	$k_t =$	4
6	4	6	13	6	0	0.000926	$k_{ord} =$	2
7	5	12	25	12	0	0.001758	$N_0 =$	1
8	6	22	47	22	0	0.003337	$G =$	1.5
...	...	...	...	...	...	...		
24	22	0	7407	6667	6667	1.000000		
25	23	0	7407	6667	6667	1.000000		

Figure 7: Fragment of the spreadsheet with repeat knowers ($S = 10\,000$, $G = 1.5$).

Computational experiment

Let $S = 10\,000$; $M = 0.6$; $B = 0.05$; $k_t = 4$; $k_{ord} = 2$; $N_0 = 1$; $G = 1.5$.

The corresponding table is shown in figure 7.

Analysing this table, we note the following:

1. At the beginning of the process, the probability of repeat knowers appearing (characterised by Q) is very small, but gradually it increases to 1 (100%), that is, all new knowers over time become repeat ones.
2. Since the increment ΔN equals the change in the number of knowers per one transmission, ΔN can be used to characterise the speed of news spreading. From the table it is evident that at the beginning of the process this speed is also very small. During the process, its speed slowly increases to a maximum and then rapidly falls to zero, remaining equal to zero for the rest of the modelling time. Simultaneously, the total number N_j of knowers of the news ceases to change (increase).
3. Let us display figure 8 – graphs of the dependencies $\Delta N = \Delta N(j)$ and $N = N(j)$ – and verify the above.

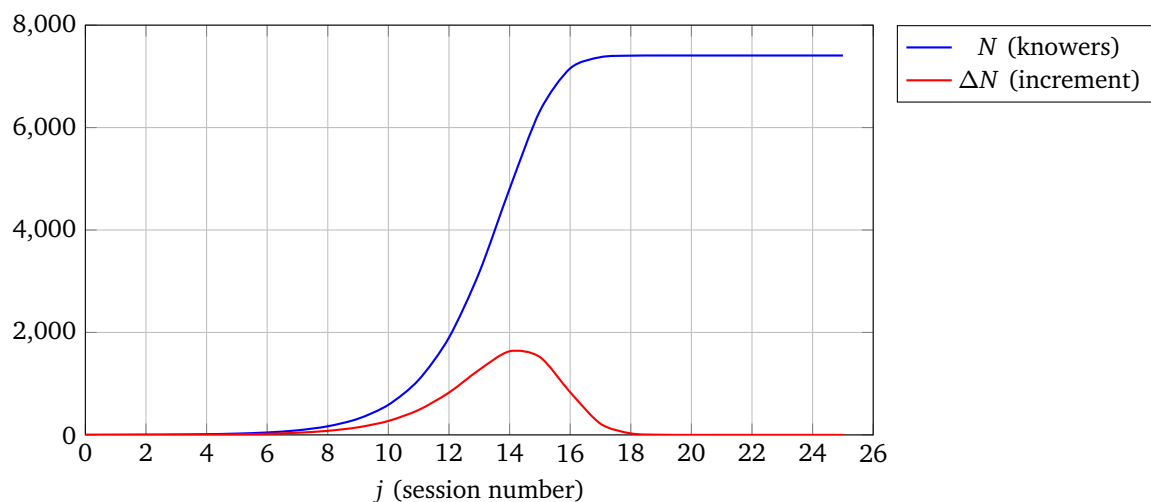


Figure 8: Graphs of $\Delta N = \Delta N(j)$ and $N = N(j)$ for the model with repeat knowers ($S = 10\,000$, $G = 1.5$).

4. A noteworthy and unexpected result stands out:

$$N_{j \max} < S.$$

This means that with the adopted input data, part of the population may remain uninformed.

Task. Estimate the size of this uninformed portion $S - N_{j \max}$ depending on the numerical values of various parameters and, first and foremost, the dependence of $N_{j \max}$ on the proportionality coefficient G in expression (6), about which it was stated that its value is selected experimentally. Take, for example, $G = 1$ (and then $G = 0.5$), leaving the remaining data at their previous values.

Exercise

1. What is the unit of rumour spreading speed according to item 2?
2. Why does $N_{j \max}$ increase (decrease)?
3. Select such a value of G that would ensure the informativeness of the entire population.
4. At what value of G will the news not spread at all?
5. What role does the coefficient B play in this version of the model?
6. If one restores the data according to item 1 and changes M – the relative share (percentage) of silent ones, for example, decreases it by 5%, then:
 - a) how should this affect the size of the uninformed portion of the population?
 - b) how does such a replacement generally affect the course of the process?
7. Repeat item 6 in turn for other parameters and comment on the results.
8. Return to item 2.2 in that part where experimentation with parameters k_t and k_{ord} is proposed. Give them the values $k_t = 7$, $k_{ord} = 5$. The table takes on an unusual appearance:
 - in consecutive adjacent cells of the column for ΔN , positive and negative values with large differences appear, constantly alternating;
 - in rows corresponding to negative numbers, the parameter $Q > 1$.

Recall that while performing previous experiments, similar situations already arose before you. Any attempts to find a logical justification here are doomed to failure in advance.

In computer modelling, such situations are well known. In the overwhelming majority of cases, and in particular when the model itself does not, by its logic, presuppose oscillatory processes, the reason for the observed significant oscillations of variables (moreover, with sign changes) is always the same: *at certain values of input data, the model (computational algorithm) loses stability*. This is one of the important features of any more or less complex model. As for the contents of the table itself, the values of variables in it are now merely the results of computations that in no way reflect the real course of the process. More details about model stability will be discussed in the next study.

Conclusions

1. The most essential quality of the model version considered is that in it, at last, there appears the long-awaited limitation on the number of knowers. Moreover, the fundamental point here is the fact that this limitation follows from the model itself, and not as a result of some additional artificial measures, as was the case in the two previous versions.
2. The computational experiment in task 4 of the Exercise allows one to conclude that the coefficient B has the meaning of a kind of regulator of the speed of the rumour spreading process.
3. An important result is obtained in item 2.3, namely the concept of model stability: *at certain values of input data, the model may lose stability, and the numerical values of variables in such cases are merely the results of computations by a corrupted algorithm and in no way reflect the real course of the process*.
4. The first signs (in medical language – symptoms) of loss of model stability are oscillations with large amplitude and sign changes of variable values in adjacent cells of some column. However, in such cases one must be certain that the observed oscillations are not a consequence of the internal regularities of the model itself, that is, that they do not follow from the nature of the object being investigated. In such situations, only personal experience of working with mathematical computer models can most often help the researcher establish the true cause of the noted oscillations.

3.4. Simplest epidemic model

We remind the reader that this problem is proposed for the purpose of consolidating the skills acquired in working with the rumour propagation problem. However, here only textual commentaries will be provided, and sample tables and graphs will be absent. You will have to perform this work independently, using the solution of the previous problem.

Problem formulation

Throughout many years of human existence, numerous lives have been claimed as a result of various epidemics. Plague, cholera, smallpox, influenza, and certain other diseases have frequently affected large masses of people.

In order to successfully combat epidemics and take necessary sanitary measures in a timely manner (quarantines, vaccinations, etc.), it is necessary to have an idea of their effectiveness. Comparing the effectiveness of these measures is possible provided there is the opportunity to determine how any of them affects the course of the epidemic, that is, how the number of illnesses will change over time. Thus, in a natural way, the need arises to create a mathematical model that would help in studying the phenomenon. *The goal of modelling will be to predict the number of sick individuals at any moment in time, given that their number at the initial moment is known.*

It is clear that a more or less realistic epidemic model must take into account the influence of many different factors, including in particular:

- regularities governing the mechanisms of bacterial cell development;
- availability of means to combat the disease;
- the degree of susceptibility to infection of individual people;
- the probability of contacts between carriers of the infection and healthy people, and much else.

That is, the epidemic model under discussion would need to be based on knowledge from various sciences (microbiology, pharmacology, medicine, sanitation, probability theory and mathematical statistics, social psychology, etc.). However, we are unable to account for the factors noted above, or even a part of them. The situation proves too complex for us, and therefore it is advisable to adopt a number of assumptions aimed at its simplification.

So as not to discourage our first steps in modelling, let us note that the disease being spread is not fatally dangerous.

Creating the model: conceptual model

Suppose that at a certain moment, several sick persons – carriers of the infection – enter a group of healthy people. Clearly, the overall numerical composition of this group (healthy + sick) may change over time, but then the study of changes in this numerical composition alone would represent an independent and not very simple task. To simplify, let us adopt a number of assumptions.

Assumption 1. During the epidemic, no quantitative changes occur in the group: there are no deaths, no births, no isolation; no one arrives and no one leaves. For not too large groups and not too lengthy time intervals, this is acceptable.

Concerning the mechanism of infection transmission, let us adopt

Assumption 2. A healthy person becomes infected and becomes a source of infection upon meeting a sick person.

In this case, we are neglecting the duration of the latent (hidden) period of the disease compared with the duration of the epidemic. Medicine is familiar with such diseases. From relevant reference books and conversations with specialists one can learn that, for example, for influenza and cholera this period is several hours, and for diphtheria, plague, and whooping cough – approximately one day or somewhat more.

Assumption 3. One sick person can meet only one healthy person (group meetings do not occur).

Assumption 4. Among the measures to combat the disease, we choose the simplest one: nothing is done in this direction. Such actions may not be humane, but after all we are dealing only with the simplest model, and the disease is not so very dangerous.

Creating the model: mathematical model

Let us divide the entire “observation” time into sufficiently small equal intervals Δt (small compared with the duration of the epidemic).

Proceeding to formalisation, let us introduce the notation:

i – the sequential number of the interval Δt ;

M_i, N_i – respectively the number of sick and healthy people at an arbitrary time interval with number i .

At the beginning of observation $i = 0$, hence $M_i = M_0, N_i = N_0$.

According to Assumption 1, at an arbitrary moment in time

$$M_i + N_i = M_0 + N_0. \quad (1')$$

Let us note here that M_i and N_i are variable quantities; over time they monotonically change in opposite directions: as M_i increases, N_i decreases.

Assumption 5. The increment of the number of sick people ΔM_i is proportional to the duration of the time interval Δt and the number of meetings of healthy people with sick ones during this interval.

Assumption 6. The number of meetings is proportional both to the number of sick (M_i) and to the number of healthy (N_i) people at the beginning of the i -th interval Δt . In that case, the number of meetings is proportional to the product of M_i and N_i .

On the basis of Assumptions 5 and 6 we can write:

$$\Delta M = k \cdot M_i \cdot N_i \cdot \Delta t, \quad (2')$$

where $0 < k < 1$ is a proportionality coefficient that we shall call the *infection coefficient*. This coefficient has a statistical meaning and takes into account, first and foremost, such factors as the probability of meetings and the susceptibility to the disease of an individual person. Let us note that different individuals in the group may exhibit different susceptibility to the disease, as a result of which as the epidemic progresses, the most susceptible will fall ill first, and at the end the most resistant will remain. Simultaneously, the probability of meetings will decrease. This means that the numerical value of k can be taken as constant.

Assumption 7. Let the infection coefficient k be constant over time.

From (1') we have:

$$N_i = M_0 + N_0 - M_i, \quad (3')$$

and after substituting (3') into (2') we finally obtain:

$$\Delta M_i = k \cdot M_{i-1} \cdot (M_0 + N_0 - M_{i-1}) \cdot \Delta t. \quad (4')$$

We find the new number of sick M_i by adding to its previous value M_{i-1} the just-computed increment ΔM :

$$M_i = M_{i-1} + \Delta M_i. \quad (5')$$

Equations (4') and (5') are the simplified mathematical model of the epidemic.

If they are solved with respect to M_i , then we shall be able to predict the course of the epidemic.

Equation (4') is said to be written in *finite-difference form*. School mathematics does not study analytical methods for solving equations of this kind. Therefore, to achieve the final goal, we shall perform this solution by a numerical method in a spreadsheet environment. We shall display the results on screen in the form of tables and corresponding graphical dependencies of the variables ΔM_i , M_i , and N_i on time.

Algorithm for working with the model

Remark. In accordance with the meaning of the variables M_i , N_i , and ΔM_i , it is advisable to display their values in *integer format*. Let us recall that in the computer's memory, and hence in arithmetic operations, they will have *real number format*.

1. Create a table; in columns A, B, C, D in the first row, place variable names: t , ΔM , M , N . Place values of these variables in subsequent rows, starting from the second.
2. In column E ("Given:"), place the names of constants: N_0 , M_0 , k , Δt with "=" signs, and in column F – their values.
3. Fill the second row:
 $t = 0$;
 $\Delta M = 0$ (no increment yet);
 M and N by references to the corresponding cells in column F with the values of M_0 and N_0 .
4. Fill the next (third) row ($i = 1$) with formulae for computing the next values of variables:
 $t_i = t_{i-1} + \Delta t$;
 ΔM according to (4'): $\Delta M_i = k \cdot M_{i-1} \cdot (M_0 + N_0 - M_{i-1}) \cdot \Delta t$;
 M according to (5'): $M_i = M_{i-1} + \Delta M_i$;
 N by expression (3'): $N_i = M_0 + N_0 - M_i$.
5. Repeat item 4 $n = 20$ times (to fill one screen). To this end, copy the formulae of the third row into the subsequent 19 rows.

Regarding the criterion for choosing the lower bound of the interval Δt values, let us note the following. If further reduction of Δt does not lead to noticeable changes in results, then its value can be left as before. However, too small a value of Δt will lead to the necessity of increasing the number of repetitions. This, in turn, will require an increase in the volume of operational memory and will correspondingly increase the data processing time. The last remark is of a *general* character, although in this specific case (with a small number of columns and rows) the mentioned increase in time will not be noticeable.

Note. By methods of higher mathematics, one can obtain an exact analytical solution of equation (4') in the form of a complex function of time, but something else is important now. After all, the answer in modelling, as is known, does not reduce simply to a number or a formula. It is desirable to obtain a table of values of the sought quantities for various moments of time, so as to plot graphs of the corresponding dependencies from this table. The study of information presented in graphical form often proves more effective than analytical analysis of formulae. However, this by no means signifies that analytical investigations in modelling are not advisable.

Computational experiment

1. Assign the following values to the model parameters:

$$N_0 = 10\,000; M_0 = 1; k = 0.1; \Delta t = 0.1.$$

This means that at the initial moment, one sick person – a carrier of the infection – enters a group of ten thousand healthy people.

- 1.1. Prepare the necessary spreadsheet and fill 22 of its rows (one screen). From the table it is evident that already at the sixth step, the stability condition of the finite-difference equation (4') ceases to hold.
- 1.2. To rectify the situation, let us reduce the interval Δt by one order of magnitude, that is, set $\Delta t = 0.01$. Now, it seems, all is well.
2. Examining several of the last rows of the table, we see that after some time all members of the group become sick. Indeed, no new sick persons appear ($\Delta M = 0$), and the total number of sick is 10 001 according to (1'). Using this table, let us plot graphs of the dependence on time of the number of sick $M = M(t)$, the number of healthy $N = N(t)$, and also the graph $\Delta M = \Delta M(t)$.

3. Let us reduce the interval Δt by half more: $\Delta t = 0.005$. We observe a slowing of the change in ΔM and M . Thus, to complete the calculations we shall now need to bring the number of repetitions to almost 50. However, we do not discover any substantial qualitative changes in the corresponding tables or graphs. It turns out that the last reduction of Δt was not obligatory. Therefore, let us leave $\Delta t = 0.01$.
4. The values of results in the column for ΔM are exclusively positive numbers. This means that the number of sick only increases over time, which is visible from the column for M ; and in the description of the model it is precisely noted that nothing is being done to combat the disease, and hence no one recovers. Here we have established nothing new, but from this fact we have seen that our model correctly reproduces an expected result in advance. Model developers even deliberately resort to such actions, introducing input data for which predictable results should be reproduced. Let us recall that such work – *model testing* – must always be performed.
5. Let us now determine how a twofold increase in the initial number of sick will affect the development time of the epidemic, that is, with $M_0 = 2$. Would it be correct to answer that such a replacement will lead to a noticeable shortening of this time? Here we have the opportunity to test the answer on the model, setting $M_0 = 2$. The result refutes the proposed answer! The epidemic spreading time does indeed shorten, but the shortening is almost imperceptible – just one row of the table, one interval Δt . This result could have been predicted by examining the table at $M_0 = 1$: after all, $M = 2$ appears there precisely in the second row.
6. The average speed of epidemic spreading $\Delta M / \Delta t$ can be estimated by the values of the increments ΔM , since all intervals Δt are equal. Both the table and the graph $\Delta M = \Delta M(t)$ corresponding to it provide the opportunity to see that the speed of epidemic spreading is not constant. Indeed, if one examines the column for ΔM in the table, it is evident that this speed changes in a complex manner: at first it grows rapidly and then decreases. From the table it is easy to determine at which step of the process the speed reaches its maximum value. But the visual clarity that a graph is capable of providing is beyond any comparison with that of a table. If one simultaneously displays the graphs of the dependencies $\Delta M = \Delta M(t)$, $M = M(t)$, and $N = N(t)$ on screen, the information obtained from them will be more attractive and convenient for comparative analysis.
Perform such a construction and analysis of each of the graphs.

Exercise

1. How are the values of model coefficients determined, in particular, k ?
2. What is the sign that the interval Δt has been chosen too large or too small?
3. What meaning should be attributed to the appearance of negative values in the column for ΔM ?
4. Experimentally prove that the product of two bounded variables that change in opposite directions (in our model, these are the variables M and N) reaches its maximum when the variables take equal values. Use the result obtained to comment on the graph of the dependence $\Delta M = \Delta M(t)$.

Conclusions

1. With regard to the dynamics of process development, the model agrees reasonably well with data from numerous observations, from which it is well known that at the beginning of an epidemic (for animals, an *epizootic*) the number of cases sharply increases and then eventually decreases.
2. The constant numerical composition of the group noted in the model description can be regarded as a recommendation, as one of the measures to combat the epidemic through the use of quarantine. If nothing had been known about *quarantine*, we could have claimed a discovery.

3. The numerical solution of this problem can be performed with pencil and paper, but such work will be long, tedious, and tiring, and errors will certainly appear as a result. It is precisely this kind of work that should, first and foremost, be delegated to computers, which are capable of rapidly and reliably processing an enormous number of cycles.

References

- [1] Teplytskyi, I.O., 2025. Elements of computer modelling: 12 studies for beginners. Study 1: Models and modelling. *Science Education Quarterly*, 2(2), pp.126–145. Available from: <https://doi.org/10.55056/seq.963>.
- [2] Teplytskyi, I.O., 2025. Elements of computer modelling: 12 studies for beginners. Study 2: Main stages of modelling. *Science Education Quarterly*, 2(3). Available from: <https://doi.org/10.55056/seq.1072>.