Simulating Pairwise Communication for Studying Its Impact on Community Public Opinion

Grygoriy Zholtkevych[1], Olena Muradyan[2], Kostiantyn Ohulchanskyi[1], and Sofiia Shelest[1]

[1] School of Math and Comp. Sci, V.N. Karazin Kharkiv National University, 4 Svobody Sqr, Kharkiv, 61022, Ukraine; http://math.univer.kharkov.ua
g.zholtkevych@karazin.ua;
spaceiseternal,soniashelest@gmail.com;
[2] School of Sociology, V.N. Karazin Kharkiv National University, 4 Svobody Sqr, Kharkiv, 61022, Ukraine; http://sociology.karazin.ua
o.s.muradyan@karazin.ua

Abstract. Communication on the fly made possible by modern information and communication technology is a characteristic feature of modern society. This style of communication significantly changed social life and provided new ways of public opinion formation. New social practice poses new challenges for specialists studying social phenomena. One of these problems is the problem of the reliability of public opinion measurements, which are always based on indirect assessments. Unfortunately, indirect assessments depend essentially on the suggestions accepted during their development. The marked above changes raised by the technological progress require the new mathematical suggestions lied in the base of public opinion measurements. This paper is to draw attention to this situation and to begin the movement toward the rigorous theory of public opinion measurements basing on social phenomena mathematical models of adequate to features of modern communication processes. It seems that authors' first results are consistent to hypotheses of a number of sociologists working in this area.

Keywords: public opinion· behaviour· microstate· macrostate· communication rate· indirect measurement· simulation

1 Introduction

The modern information and communication technology (ICT) has essentially changed communication processes between communities and their members, generated new social phenomena. Social networks are of these phenomena. Their emergence and rapid development have markedly changed the character of communication streams.

In this context, an important point is the effect of a combination of social networks and personal telecommunication microelectronic devices (tablets, smartphones). This effect significantly speeds up communication, reduces in time and

makes it discrete and as non-linear as possible. The cognitive and motivational results of such communication are still unknown to sociologists. Nevertheless, it can be argued that such discrete and nonlinear communication significantly diversifies the sources of potential influence on the opinion formation of the of communication participants, and at the same time increases the frequency of such impact, potentially having a different result. Public opinion formation becomes the process difficult for predicting because we need to take into account a significant number of factors than before. At the same time, the effect of these factors due to the specificity of the new segment of the communicative space (nonlinearity, discreteness and high communication rate) is characterised by more and more complex interconnections.

The reliable sociological information about public opinion is especially in demand in modern society. The reason for this is not to reinforce the need to effectively predict changes in public opinion (this need has traditionally been high since the fifties of the twentieth century). The reason is associated with an increase in such threats to social stability and security as information terrorism, manipulative technologies affecting society, and the active use of fake-news in the political sphere. Today public opinion needs not only to predict but also to protect against the above threats. This, in turn, requires the development of the technique of sociological measurement based on the dependable quantitative theory. This necessity is also caused by the high cost of the evidential public opinion measurements. Improving public opinion measurements by the way of increasing the frequency of spot measurements as to record in time all possible deviations and shifts in public opinion seems to be of little prospect due to the high cost of such measurements and their technical complexity.

Based on the foregoing, it can be said that developing a theory of measuring public opinion to substantiate and improve polling tools aimed at making it possible to grasp the peculiarities of the formation of public opinion in modern conditions is a relevant challenge not only for social science but also for mathematics and computer science.

We believe that simulation of the public opinion formation within the modern information and communication environment is the origin point for this theory.

This paper is our attempt to attract the attention of researchers in the fields of ICT, mathematics and sociology to this challenge.

It is needed to stress that the idea to use mathematical modelling as a tool for studying social processes is not novel. In the middle of the twentieth century, the use of relational and statistical models for understanding social processes was proposed by a number of scientists (N. Rashevsky [13], A. Rapoport [12] and other). The idea was further developed with the inception of network science (see [1]).

Today there are a lot of works devoted to simulation of social network and studying social dynamics based on the corresponding methods. But we could not find a paper that would present a simulation model for studying a public opinion measurement. This is what caused our research.

2   Model of Communication in Network

In this paper, we consider the special class of homogeneous multicomponent discrete-time dynamical systems, whose components interact only pairwise via a network of channels. The components of such a system are entities for modelling members of the community being studied, and channels are entities for modelling stable pairwise communications between the members of this community.

2.1   Modelling Assumptions

The goal of this subsection is to formulate explicitly the modelling assumptions.

Assumption 1. A network being simulated is a multicomponent dynamical system of the discrete time with a constant set of components called members. Some pairs of components communicate stable and, in this case, we say that there is an information channel between members of such a pair.

Assumption 2. The property of an information channel to be active is a random Boolean variable. Moreover, for different channels, the corresponding variables are independent.

Assumption 3. Each system component estimates the claim being in the focus of community interests using an element of the set $\mathbf{C} = \{\mathsf{RED}, \mathsf{GREEN}, \mathsf{BLUE}\}$ corresponding to either negative, or neutral, or positive estimation respectively. The mapping associating a claim estimation with each system components is below called a microstate of the system.

Assumption 4. The personal estimations of a member of a community (a system microstate) is not interest and is considered as not available for the direct observation. Only the occupacy measure of elements of $\mathbf{C}$ is available for the direct observation. The value of this measure for an element of $\mathbf{C}$ is a ratio between the number of community members that have the corresponding opinion and the total number of community members. This measure is considered below as a macrostate of the system.

Assumption 5. At each time-point, a member of the community participates at most in one communication.

2.2   Specification of Network Model

Taking into account Assumption 1, an undirected simple finite graph $\mathbf{G} = (\mathbf{N}, \mathbf{E})$ is the most natural mathematical structure for modelling pairwise communications in a community. The node set $\mathbf{N}$ of the graph models members of a community, and the edge set $\mathbf{E}$ of the graph models stable information channels between ones.

Assumption 2 can be ensured with associating a random Boolean variable $\mathsf{activated}_e$ with each edge $e \in \mathbf{E}$. To do this it is sufficient to specify a function

activation_rate : $\mathbf{E} \rightarrow [0,1]$ and to think about its value activation_rate(e) as about the activation probability of the channel corresponding the edge e. In other words, we set activation_rate$(e) = \Pr(\mathsf{activated}_e = \mathbf{true})$. Thus, we come to the following definition.

Definition 1. *A network model is a triple* $\langle \mathbf{N}, \mathbf{E}, \mathsf{activation\_rate} \rangle$ *where* $\mathbf{N}$ *and* $\mathbf{E}$ *are respectively the sets of nodes and edges of some undirected simple finite graph* $\mathbf{G} = (\mathbf{N}, \mathbf{E})$ *and* activation_rate $: \mathbf{E} \rightarrow [0,1]$ *is the channel activation rate function.*

Assumption 3 causes the following definition of a microstate and micro-dynamics for the system class being studied.

Definition 2. *A node colouring of the graph* $\mathbf{G}$ *in accordance with the colour set* $\mathbf{C}$ *is a microstate of the system.*
Thereby, the system micro-dynamics is a discrete-time stochastic process explaining the observed sequences of system microstates.

Assumption 4 leads us toward the concepts of a macrostate and macro-dynamics.

Definition 3. *Let* $c : \mathbf{N} \rightarrow \mathbf{C}$ *be a microstate of the system then a function* $\overline{c} : \mathbf{C} \rightarrow [0,1]$ *is the macrostate corresponding to c if for each* $x \in \mathbf{C}$*, it is defined as follows*[3]

$$\overline{c}(x) = \frac{1}{|\mathbf{N}|} \cdot \sum_{n \in \mathbf{N}} [c(n) = x].$$

Thereby, the system macro-dynamics is a discrete-time stochastic process explaining the observed sequences of system macrostates.

2.3 Simulation Framework Concept

Based on the above assumptions and definitions, a prototype framework has developed for simulation of the community dynamics with various kinds of pairwise communications. The general specification of a simulation process is presented as a UML activity diagram in Fig. 1. For the realisation of this general specification, the language Python 3 [11] and library NetworkX 2.2 [9] have been used.

To construct a framework providing the presented simulation process we propose the conceptual model shown as a UML class diagram in Fig. 2. This model based on an undirected simple graph whose nodes are instances of the class Node and the edges are instances of the class Edge. The attribute estimation of the class Node is intended for saving the current value of a microstate for the corresponding node. The association state gives access to the internal description of a node state. This description is abstract on the framework level. Similarly, the attribute activation_rate of the class Edge is intended for saving the value activation_rate(e) for the Edge-instance that models edge e.

---

[3] In the formula, the Iverson bracket is used (see, for example, [5]). The value of $[c(n) = x]$ equals 1 if $c(n) = x$, and otherwise it equals 0.
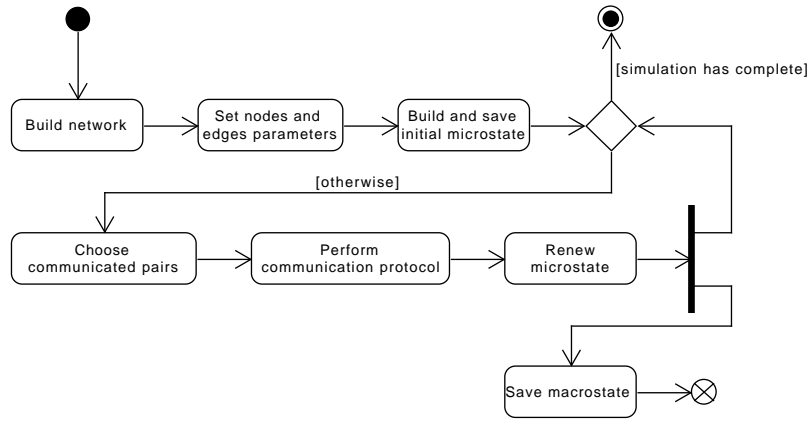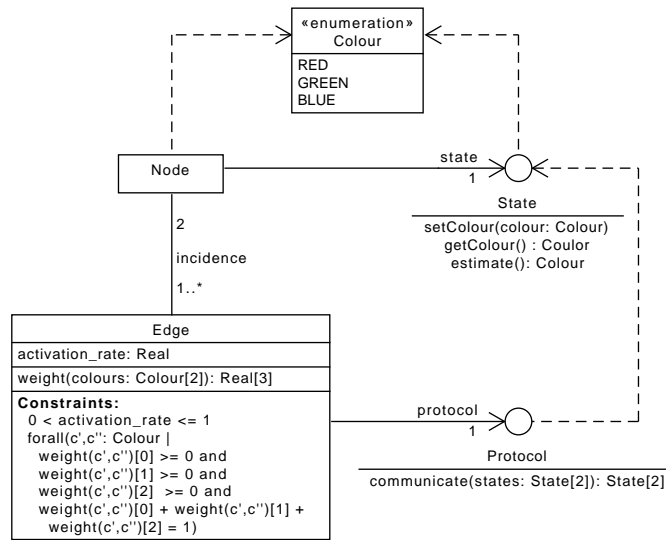
Fig. 1. General specification of a simulation process

Fig. 2. Conceptual model of the simulated net

2.4   Pairwise Communication Model

Above we were focused on modelling the structure of a network, and in this subsection, we pass to modelling the interaction (or the communication, in the case of social network) between nodes of the network. The association between instances of the class Edge and abstract entities classified as Protocol is foreseen for providing the specification of such interaction (see Fig. 2).

Taking into account that Assumption 5 is accepted we need some method to form the set of interacting pairs of nodes. We propose to use the method specified by Algorithm 1.

Algorithm 1: Method for forming communicating pairs

Data: a simple undirected graph $\mathbf{G} = (\mathbf{N}, \mathbf{E})$
Result: the subset SELECTED of E representing communicating pairs

```
  /* initialise the target set and auxiliary sets                  */
1 SELECTED := ∅; AVAILABLE := ∅; FORBIDDEN := ∅;
  /* activate communication channels                               */
2 foreach e ∈ E do
3    │ choose randomly True or False with probabilities a(e) and 1 − a(e)
     │   respectively;
4    │ if True is selected then
5    │    │ add e into AVAILABLE
6    │ else
7    │    │ add e into FORBIDDEN
8    │ end
9 end
  /* form communicating pairs                                      */
10 while AVAILABLE ≠ ∅ do
11    │ choose randomly an element e ∈ AVAILABLE in accordance with uniform
      │   distribution on AVAILABLE;
12    │ delete e from AVAILABLE;
13    │ if for some e′ ∈ SELECTED, e and e′ are incindent then
14    │    │ add e into FORBIDDEN
15    │ else
16    │    │ add e into SELECTED
17    │ end
18 end
```

The following proposition establishes properties of the method.

Proposition 1. *The method presented by Algorithm 1 has properties*

1. *a computation with respect to Algorithm 1 is halted for any input data after a finite number of steps;*
2. *after halting a computation with respect to Algorithm 1, sets* SELECTED *and* FORBIDDEN *are disjoint;*

3. *after halting a computation with respect to Algorithm 1, set* SELECTED *does not contain incident edges;*
4. *adding to the set* SELECTED *an edge added to the set* FORBIDDEN *in loop 10–18 violates the property claimed in item 3.*

Proof. The first item of the proposition is true because of the set AVAILABLE decreases (see, line 12 of Algorithm 1) after each iteration of loop 10–18.

The validity of the second item of the proposition is ensured by branching 13–17.

The validity of the third item of the proposition is ensured by line 14.

The validity of the fourth item of the proposition is ensured by branching 13–17. $\square$

We suggest that any communication protocol can be represented by the UML sequence diagram as in Fig. 3.
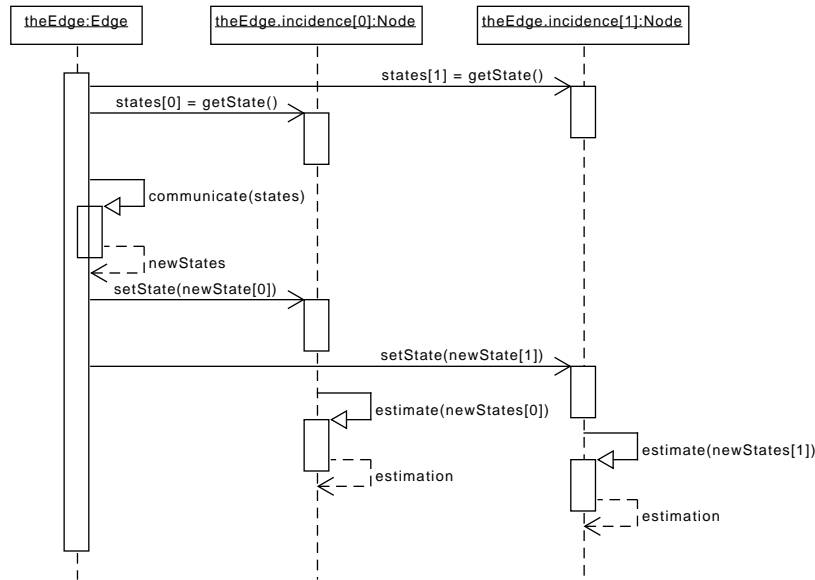

Fig. 3. Model of a pairwise communication protocol

Finally, the method estimate(state: State) of the abstract entity State (see Fig. 2) is intended to renew the current microstate.

3 Computational Case Studies

In this section, we present and discuss the results of simulation for four kinds of systems: models A-IR and B-IR, which called below as models of components

with an instant response, and models A-LR and B-LR, which called below as models of components with a lazy response.

3.1 Realisation of the Method communicate(...)

The above classification of the models being studied is based on the general scheme of the interaction process modelled by the method communicate(...) of the abstract entity Protocol.

We assume that the communication corresponding to edge $e \in \mathbf{E}$ is modelled by the weight function w_e on $\mathbf{C} \times \mathbf{C}$ and taking random value $w_e(colour', colour'')$ in the following outcome set $\{\mathsf{nobody}, \mathsf{first}, \mathsf{second}\}$. The outcome is interpreted as follows

- $w_e(colour', colour'') = \mathsf{nobody}$ means that participants of the communication preserve their opinions;
- $w_e(colour', colour'') = \mathsf{first}$ means that the first participant of the communication preserves his opinion, but the second one does not preserve;
- $w_e(colour', colour'') = \mathsf{second}$ means that the second participant of the communication preserves his opinion, but the first one does not preserve.

Based on this assumption, we propose to use the following abstraction specified by Algorithm 2.

Algorithm 2: The scheme of the method communicate(...)

Data: an edge $e \in \mathbf{E}$, the weight function w_e corresponding e
Result: the pair of new node states (newFirstState, newSecondState)

1 firstState := e.incidence[0];
2 secondState := e.incidence[1];
3 choose randomly outcome from $\{\mathsf{nobody}, \mathsf{first}, \mathsf{second}\}$ in accordance with the distribution w_e(firstState.getColour(), secondState.getColour());
4 **if** outcome = nobody **then**
5 | newFirstState = firstState;
6 | newSecondState = secondState
7 **else if** outcome = first **then**
8 | newFirstState = firstState;
9 | create newSecondState in accordance with a concrete algorithm
10 **else** /* outcome = second   */
11 | create newFirstState in accordance with a concrete algorithm;
12 | newSecondState = secondState
13 **end**
14 **return** (newFirstState, newSecondState)

Remark 1. Note that everywhere below we use the weight function defined as follows

1. $w_e(c, c) = \{\text{nobody} = 1.0,\ \text{first} = 0.0,\ \text{second} = 0.0\}$ for any $c \in \mathbf{C}$;
2. $w_e(c', c'') = w_e(c'', c')$ for all $c', c'' \in \mathbf{C}$;
3. $w_e(\text{GREEN}, c)[\text{nobody}] = 0.0$,
   $w_e(\text{GREEN}, c)[\text{first}] = 0.1$, and
   $w_e(\text{GREEN}, c)[\text{second}] = 0.45$ for any $c \in \{\text{first}, \text{second}\}$;
4. $w_e(\text{RED}, \text{BLUE})[\text{nobody}] = 0.1$,
   $w_e(\text{RED}, \text{BLUE})[\text{first}] = 0.45$, and
   $w_e(\text{RED}, \text{BLUE})[\text{second}] = 0.45$.

3.2 Systems of Components with Instant Response

The model of a system of components with instant response (below IR-model) is based on the following model of a state called by SimpleState (see Fig. 4).
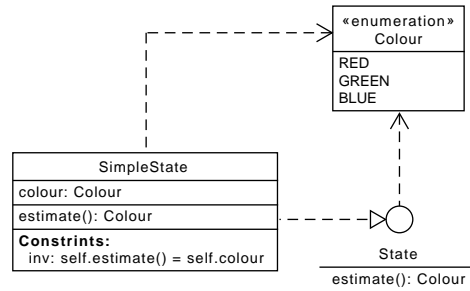

Fig. 4. Model of a simple state

The IR-model realises items 9 and 11 of Algorithm 2 as follows

if outcome $=$ nobody then

$$
\begin{aligned}
\text{newFirstState} \quad &= \text{firstState} \\
\text{newSecondState} &= \text{secondState}
\end{aligned}
$$

if outcome $=$ first then

$$
\begin{aligned}
\text{newFirstState} \quad &= \text{firstState} \\
\text{newSecondState} &= \text{firstState}
\end{aligned}
$$

if outcome $=$ second then

$$
\begin{aligned}
\text{newFirstState} \quad &= \text{secondState} \\
\text{newSecondState} &= \text{secondState}
\end{aligned}
$$

Simulation Experiment for the IR-model. The simulation experiment was carried out at the initial macrostate defined as follows $\bar{c}_0(\mathsf{RED}) = 0.1$, $\bar{c}_0(\mathsf{GREEN}) = 0.8$, and $\bar{c}_0(\mathsf{BLUE}) = 0.1$. The typical simulation results are shown in Fig. 5.
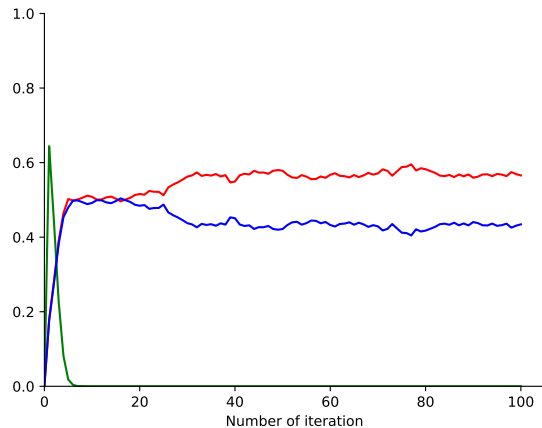

Fig. 5. A typical behaviour of the IR-model

Error Estimation for the IR-Model. We assume that the measurement of the system is performed sequentially by observing a fixed number of system components. Thus, the measurement rate depends on the number of observed components in one step. More precisely, our assumption is that the measurement procedure under our study is sequential and represented by the Algorithm 3.

We estimate the measurement error by using Kullback-Leibler divergence [7, 6, 3] $D(\bar{c} \,||\, \bar{c}_*)$ where $\bar{c}$ is the real system macrostate and $\bar{c}_*$ is the measured system macrostste at the end of simulation.

Remind that Kullback-Leibler divergence D is computed by the formula

$$D(\bar{c} \,||\, \bar{c}_*) = \sum_{c \in \mathbf{C}} \bar{c}(c) \cdot \log_2 \frac{\bar{c}(c)}{\bar{c}_*(c)}$$

and estimates the minimal information quantity needed to correct an error.

As mentioned above, the measurement speed depends on the number of k nodes observed during one simulation cycle. A small value of k corresponds to a slow measurement and a big value of k corresponds to a fast one. In Fig. 6, the dynamics of error estimation for slow (the blue curve with $k = 20$) and fast (the green curve with $k = 250$) measurements are presented.

Algorithm 3: Measurement procedure

Data: a model of a system, a number k of nodes observed per one simulation cycle

Result: the measured macrostate $\overline{c}_*$

1 $N[\mathsf{RED}] = N[\mathsf{GREEN}] = N[\mathsf{BLUE}] = 0$;

2 foreach *simulation cycle* **do**

3   choose randomly k nodes from the nodes not chosen yet;

4   increase each $N[\mathsf{RED}]$, $N[\mathsf{GREEN}]$ and $N[\mathsf{BLUE}]$ by the number of nodes from the sample correspondingly coloured

5 end

6 $N = N[\mathsf{RED}] + N[\mathsf{GREEN}] + N[\mathsf{BLUE}]$;

7 $\overline{c}_*(\mathsf{RED}) = N[\mathsf{RED}]/N$;

8 $\overline{c}_*(\mathsf{GREEN}) = N[\mathsf{GREEN}]/N$;

9 $\overline{c}_*(\mathsf{BLUE}) = N[\mathsf{BLUE}]/N$;
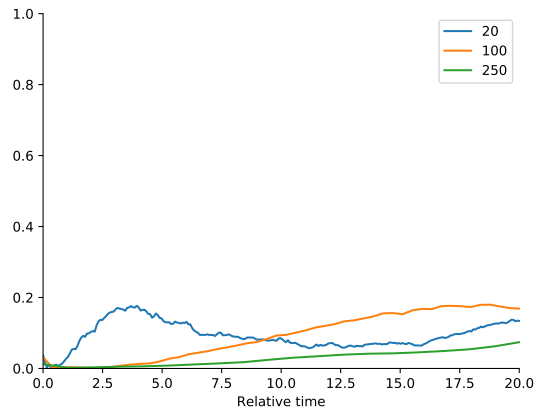
10 return $\overline{c}_*$

Fig. 6. Error for the measurement with $k = 20$, for the measurement with $k = 100$, and for the measurement with $k = 250$

Looking in Fig. 6 one can see that the error estimation increases with increasing of the measurement rate. This means that there exists perhaps some low bound for the precision of a measurement.

3.3 Systems of Components with Lazy Response

The LR-model is based on the following model of a state called by LazyState (see Fig. 7).
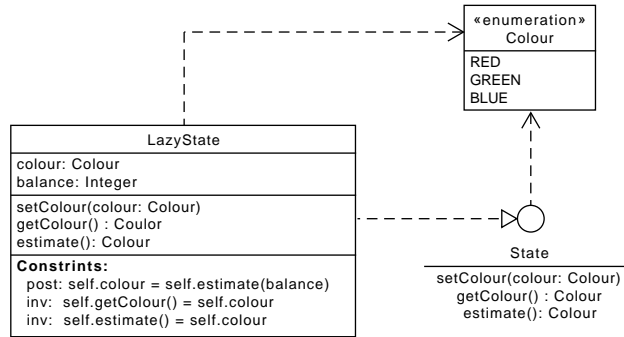

Fig. 7. Model of a lazy state

Unlike the previous model, the model considered in this subsection is more inertial. This is provided by the method estimate(), which uses the function $P_m(x)$, and the field balance, which equals the difference between BLUE-arguments and RED-arguments (see Fig. 7).

The function $P_m(x)$ is defined as

$$
P_m(x) = \begin{cases} \dfrac{x^3}{m^3}\left(1 - \dfrac{x}{2m}\right) & \text{if } 0 \le x < m \\ \dfrac{1}{2} + \dfrac{1}{\pi}\arctan\dfrac{\pi(x-m)}{m} & \text{if } x \ge m \end{cases}
$$

This function provides model inertness. Its value equals the probability that the corresponding system node is not green. We assume that the current balance of the node determines this probability.

The LR-model realises items 9 and 11 of Algorithm 2 as follows

if outcome = nobody then

$$
\begin{aligned}
\text{newFirstState} &= \text{firstState} \\
\text{newSecondState} &= \text{secondState}
\end{aligned}
$$

if outcome = first then

$$newFirstState = firstState$$
if firstState.colour = RED then
$$newSecondState.balance = secondState.balance - 1$$
if firstState.colour = GREEN then
$$newSecondState.balance = secondState.balance$$
if firstState.colour = BLUE then
$$newSecondState.balance = secondState.balance + 1$$

if outcome = second then

$$newSecondState = secondState$$
if secondState.colour = RED then
$$newFirstState.balance = firstState.balance - 1$$
if secondState.colour = GREEN then
$$newFirstState.balance = firstState.balance$$
if secondState.colour = BLUE then
$$newFirstState.balance = firstState.balance + 1$$

The positive parameter m controls the system inertia and in a certain sense can be considered as a mass. This interpretation is illustrated by Fig 8.

We should mark that the character of the measurement error behaviour is similar to one for the IR-model. This is a reason to omit the corresponding illustrating figure.

4   Conclusion

Thus, the paper has proposed a framework for simulating pair-chatting in communities. The simulation results show that our fears associated with a fundamental change in social behaviour caused by the widespread use of modern information and communication technologies are not groundless. Moreover, these changes have led to a violation of the basic assumptions on which the mathematics of sociological measurements is based. The main argument in favour of such a conclusion is the observable fact, saying for the existence of a positive lower bound for measurement errors. The mention of this effect demonstrated by simulation modelling was described in the works of sociologists devoted to the survey method. Their reasoning is informal and far from mathematical ones. In the context of this reasoning, sociologists noted the existence of distortion effects always present in such measurements. In the context of this reasoning, sociologists noted the existence of distortion effects always present in such measurements. One can mention, for example, the book of Walter Lippmann [8] and the article of Pierre Bourdieu [2]. One can also refer to the Noelle-Neumann hypothesis [10] about the spiral of silence, which illustrates the contradiction of the internal processes of the functioning of public opinion and the problems of understanding and overcoming this contradiction by sociological means.
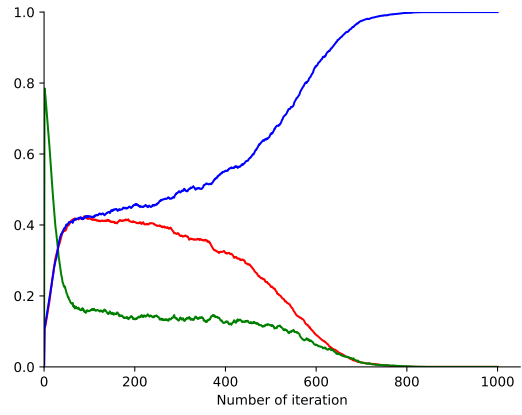
a) $m = 2$
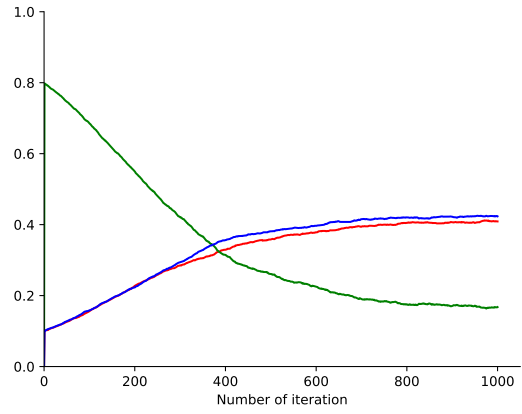
b) $m = 50$

Fig. 8. Behaviours of LR Model

In the case, if this hypothesis is confirmed, we will have to admit that the assumption of complete observability [4, p. 14] is wrong for intensively communicating communities. In other words, for studying such communities we need to use models similar to rather quantum than classical models of physical systems. Of course, this does not mean that mathematics of quantum theory is adequate for describing dynamics of intensively communicating communities. Hence, the challenge to find the adequate mathematical language for studying this class of systems.

Summing up our discussion, we can formulate the following problems for the top-priority research

1. conduct a detailed study of the dependence of the behaviour of the LR-model on the parameter m;
2. establish the dependence of the measurement error on the rate of this measurement;
3. generalise the obtained results for more complicated than pairwise communications;
4. build a simulation model for communities exposed to external influences;
5. establish the character of the dependencies between parameters of the external influence and the system behaviour;
6. find out whether the community exposed to external influences is a system managed by these influences.

If all these studies give a positive result then the problem to ensure certain community behaviour in the presence of limited resources that provide external influence on the system can be set.

References

1. Barabási, A.: Network science. Cambridge University Press (2018)
2. Bourdieu, P.: The three forms of theoretical knowledge. Social Science Information **12**(1), 53–80 (1973)
3. Cover, T., Thomas, J.: Elements of Information Theory. Wiley-Interscience, 2nd edn. (2006)
4. Holevo, A.: Probabilistic and Statistical Aspects of Quantum Theory. Scuola Normale Superiore Pisa (2011)
5. Knuth, D.: Two notes on notation. American Mathematical Monthly **99**(5), 403–422 (1992)
6. Kullback, S.: Information Theory and Statistics. John Wiley & Sons (1959)
7. Kullback, S., Leibler, R.: On information and sufficiency. Annals of Mathematical Statistics **22**(1), 79–86 (1951)
8. Lippmann, W.: Public Opinion. Harcourt, Brace and Company, New York (1922)
9. Networkx, https://networkx.github.io/, (accessed 30.12.2019)
10. Noelle-Neumann, E.: The theory of public opinion: The concept of the spiral of silence. In: Anderson, J. (ed.) Communication yearbook, vol. 14, pp. 256–308. Sage Publications, Inc., Thousand Oaks, CA, US (1991)
11. Python, https://www.python.org/, (accessed 25.12.2018)

12. Rapoport, A.: Contributions to the theory of random and biased nets. Bulletin of Mathematical Biophysics **19**, 257–277 (1957)
13. Rashevsky, N.: Mathematical Theory of Human Relations: An Approach to Mathematical Biology of Social Phenomena. Principia Press, Bloomington, 2nd edn. (1949)