Information Diffusion Prediction in Mobile Social Networks with Hydrodynamic Model

[1,2]Ying Hu, [1,2]Min Chen
[1] School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China
[2] Embedded and Pervasive Computing Lab, Huazhong University of Science and Technology, Wuhan 430074, China
E-mail: yinghu@hust.edu.cn, minchen@ieee.org
corresponding author: Min Chen

***Abstract*—Mobile social networks have gained tremendous popularity among hundreds of millions of Internet users due to their fast information spreading and strong inter-person influence. However, the high complexity of social interactions and the intrinsic dynamics of mobile social networks make it challenging to model the spreading mechanism delicately and enable precise prediction of information diffusion. In this paper, we are the first to exploit physical hydrodynamics to model the process of information diffusion in mobile social networks. With our proposed hydrodynamic information diffusion prediction model (hydro-IDP), we can accurately capture the information diffusion process from both temporal and spatial perspectives, and shed more light on the information spreading characteristics (e.g., information popularity, user influence, social platform diffusivity, etc.). We also conduct a large-scale trace-driven validation to verify the accuracy of our model. The results show that the hydro-IDP model is competent to characterize and predict the process of information propagation in mobile social networks.**

I. Introduction

With the development of 5G mobile communication systems [1] [2], everything and everyone have been connected to form a new network, i.e., social networks. In recent years, the mobile social networks (MSNs) (e.g., Facebook, Twitter, Digg, etc.) have tremendously grown up, and the information in various forms, such as text, picture, audio, and video, is spread effectively through these mobile social networks. Undoubtedly, information exchange over MSNs has become more and more popular and changed people's daily life and the way they communicate with their friends and family members.

The massive scale of information contents has accelerated research on information diffusion in MSNs. These research results can help people understand the process of the information spreading better, facilitating the prediction of those social activities more effectively [3]. But it is highly challenging to analyze the mechanism of information diffusion because of the complexity of social interactions and network dynamics of MSNs.

There are two main categories of diffusion models in the prior research, i.e., the explanatory models and the predictive models [3]. The explanatory models strive to retrace the spreading path of the information in MSNs. Most of them have focused on the measurement and analysis of the network structures [4]–[7], user interactions [8]–[13], and spreading characteristics of the social media, such as empirical approaches [14] which utilize data mining [15], [16] and statistical modeling schemes [17]–[19].

The objective of predictive models is to predict how a specific diffusion process would unfold in a given network, based on the research results of the past process of the information spreading, such as the independent cascades model [18], [20], [21] and the linear threshold model which are established based on static graph structure [24]–[26]. Several studies for information diffusion on temporal pattern are based on epidemic model [27], [28] and linear influence model [5], and a few recent effects use a partial differential equation model to predict the information diffusion on both temporal and spatial dimensions [22], [23], [29].

In this work, we propose a hydrodynamic model to describe the information diffusion process in the MSNs. To the best of our knowledge, this article is the first attempt to use hydrodynamic model to study the information diffusion process.

The rest of the paper is organized as follows. In Section II, we introduce the theoretical framework for the Hydro-IDP model. In Section III, we present the spatio-temporal patterns in a real dataset collected from Sina micro-blog site, and validate the proposed hydro-IDP model in terms of predicting the information diffusion process. After that, we study the information spreading characteristics such as the information popularity, the influence of the information publisher, the website diffusivity, etc. Finally, we conclude the whole paper and discuss the future work in Section IV.

II. Framework for IDP through Hydrodynamic Model

A. Framework of the Hydro-IDP Model

The proposed Hydro-IDP model includes the following functional components: data acquisition and analysis, hydrodynamic modeling, parameter setting and the specified information diffusion prediction as a goal. In Fig.1, we provide a flow diagram to illustrate the concept of the Hydro-IDP model.

More specifically, it includes three main steps below:

1) Determine the initial conditions for the hydro-IDP model from the specified information. We should determine the initial energy density distribution and its maximum value, and the initial source radius.
2) Obtain the common parameters, such as flow velocity, step width of time and space, for the hydro-IDP model

978-1-4799-6664-6/16/$31.00 ©2016 IEEE

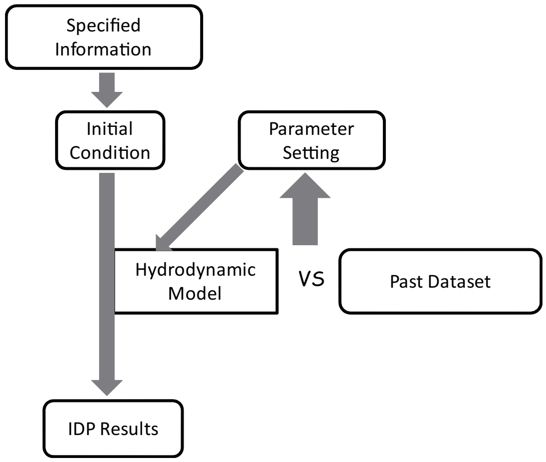

Fig. 1. The framework of Hydro-IDP.

by comparing the past datasets with the modeling results of the spatial-temporal density distribution of the influenced users for information.

3) Predict the diffusion and influence of a specified information.

Hydrodynamic frame and the determination of the common parameters are the key components of the hydro-IDP model. So we provide more detailed discussions of these components later.

B. Hydrodynamic Frame

The hydrodynamic model is based on a set of partial differential equations widely used in the physics, engineering and meteorology, etc [30]. In the ideal situation, the spatio-temporal diffusion of any flowing quantity can be modeled through hydrodynamic model with an initial condition and an equation of state, while ignoring the detail of the interior interaction. To the best of our knowledge, our paper is the first to utilize hydrodynamic model for information flow. In MSNs, the formulation of the conservation law is similar to that for spatial biology [29], [31], and one can model the evolution of the information flow with hydrodynamics by extend the discrete points of the information flow into a continuous interval.

In this work, we use a quantity of the density of the influenced users to present the evolution of the information flow and then model the information diffusion in mobile social networks through hydrodynamics. First, we give the physical meaning of the model parameters and their counterpart definition in OSNs in Table I.

With these parameters, the ideal hydrodynamic description for the system of a information diffusion is defined by local energy density conservations,

$$\partial_\mu T^{\mu\nu} = 0, \tag{1}$$

where $T^{\mu\nu} = (E+p)u^\mu u^\nu - pg^{\mu\nu}$ is the energy-momentum tensor for ideal fluid, E is the energy density and p is the pressure in the local rest frame of a fluid element moving with velocity u^μ. In our model, we define the invariant-time coordinate as $X^\mu = (t, r)$, and the metric tensor as $g^{\mu\nu} = \mathrm{diag}(1,-1,-1,-1)$ [30].

Considering the cyber-distance of friendship hops in mobile social networks and the isotropic diffusion, the hydrodynamic equation (1) can be written as below in the spherical symmetry frame

$$\begin{aligned} \partial_t E + \partial_r[(E+p)v] &= -\frac{2v}{r}(E+p), \\ \partial_t M + \partial_r(Mv+p) &= -\frac{2v}{r}M, \end{aligned} \tag{2}$$

where M is a so-called momentum density which corresponds to the energy density E.

In this paper, the algorithm leading to the solution to the hydrodynamic equations (2) is based on the Godunov method [32], which introduces finite cells and computes fluxes between cells using the Riemann problem solution for each cell boundary, and solves the Riemann problem by using the Harten-Lax-van Leer-Einfeldt (HLLE) solver [33]. To close the hydrodynamic equations (1), we simply use the equation of states of ideal flow, $p = 2/3\ E$.

III. IDP WITH HYDRODYNAMIC MODEL

A. Characteristics of information diffusion in real dataset

The dataset used in this article is collected from the most popular microblogging social network in China, i.e, Sina micro-blog. It includes some of representative video tweets which have propagated through friendship in a month. When a video tweet is published by a source user, his/her followers are able to see and repost/comment this video tweet, and so on. So the dataset provides us an opportunity to study the impact of the friendship relationship of the information spreading. It is necessary and also important to study and predict the information diffusion through social networks which are especially built on friendship pattern.

To be more specific, the datasets consist of 6500 video tweets that were propagated on Sina micro-blog during May 2012 to February 2013. We collect the user property of all repost/comment actions for each video tweet, such as the user ID, follower-counts, friendships relationship, action timestamp, etc. We consider the users who have reposted or commented the tweets as influenced users. In total, there are more than 200 million user records cast on these video tweets. These action data of the timestamps and friendship hops provide the opportunity to calculate the density of the influenced user in both temporal and spatial dimensions. In this paper we analyze all of the video tweets and demonstrate the results of three representative tweets of different following-user scales. Tweet 1 is the most popular one with 92992 followings, tweet 2 and tweet 3 have been followed 65660 and 35401 times, respectively.

Fig.2 (a) shows the distributions of the influenced user with time of three representative video tweets. We can find that 95 percent influenced users repost or comment the tweets in eight days for all three tweets. We also find from Fig.2 (b) that 98

TABLE I
MODEL PARAMETERS VS. DEFINITION IN MSNS

Symbol	Physical Meaning	Definition in OSNs
S	the centre space-time point.	the information publisher.
E (T)	the energy density (temperature) of flow.	the popularity of the information.
R	the initial source radius.	the influence of the publisher.
v	the initial flow velocity.	the diffusivity of the social platform.
r	the space distance.	the friendship hops between users.

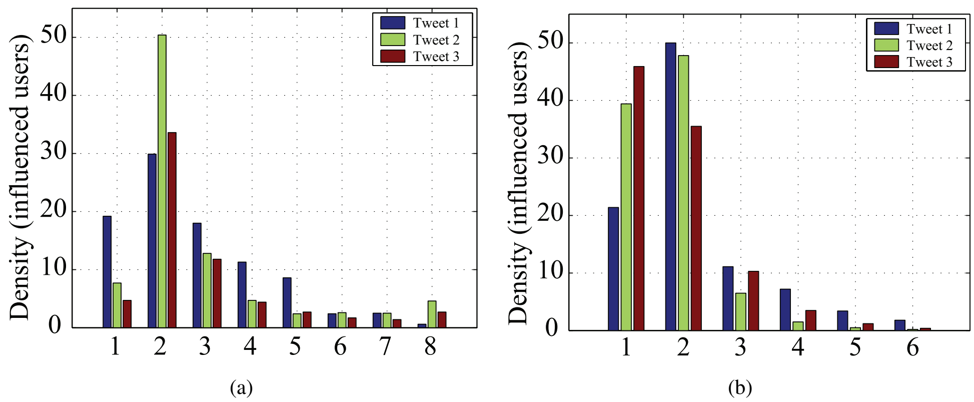

Fig. 2. (Color online) Density of the influenced users for three representative tweets with (a) days as time; (b) friendship hops as distance.

percent influenced users repost or comment the tweets in six friendship hops as distance. So we only present the results for influenced users of time 1 to 8 days and distance 1 to 5 friendship hops in the later of our work.

Fig.3 (a-c) illustrate the density of influenced users at 1 to 5 friendship hops over 8 days for three tweets which have been shown in fig.2. Each line represents the density at a certain distance.

We have also shown the density of the influenced users with time and distance for other tweets such as (b) and (c) in the Fig.3. There are only 6.7 and 3.8 percent of the total tweets, respectively. The density of influenced users of 2 friendship hops away is higher than that of users 1 friendship hops away in Fig.3 (b). This is because that there are more influential users who have reposted the tweet of 15 than the most case which less than 3, and then obviously improve the spreading of this tweet in the distance 2. In our work, we consider the user which has more than one hundred thousand followers as an influential user. In picture (c), we can find a sudden rise of the density at the seventh day (i.e., New Year's day), which further promotes the information spreading.

Our analysis results have shown that almost 87.3 percent of all the tweets in our dataset fall into the category of tweet1. The density of influenced users evolves with time and becomes stable at the fifth day. This illustrates that the tweet is no longer 'new' and has stopped its spreading process. The density of influenced users at distance 1 is higher than that of users with hops greater than 1. This indicates that the direct link of the friendship plays an important role in information diffusion.

In Fig.4, the dash lines illustrate the density of influenced users of the most popular tweet 1 from a different perspective. We give the density value with distance of friendship hops for the different time from 1 to 7 days. This picture shows the spatio-temporal distribution of the influenced users for the tweet. We can see that the density value decreased obviously with time and distance.

B. Results of hydro-IDP model

The past empirical study shows that the result of the information diffusion presents different temporal and spatial patterns with a variety of reasons, such as spreading platform, structure of the social network, the detail rapid change of the network, the social interaction, and so on. Such elements make it more challenging to analyze and predict information diffusion. In this subsection, we propose a hydro-IDP model to give the spatio-temporal evolution of tweets. By using hydrodynamics, one can model the spatio-temporal evolution of the source density distribution once one provides the initial condition and the equation of state, without considering the detail of the interaction of the users and the change of mobile social networks.

In Fig.4, the solid lines are the density of influenced users for the model results. In our model validation, we fix the flow velocity $v = 1$ for Sina micro-blog platform, and give an initial source with radius $r = 5$ which corresponds to the tweet publisher's 1.5 million followers scale. And then we adjust the initial temperature (corresponding to the initial energy density) of the source decrease linearly from maximum $T = 300$

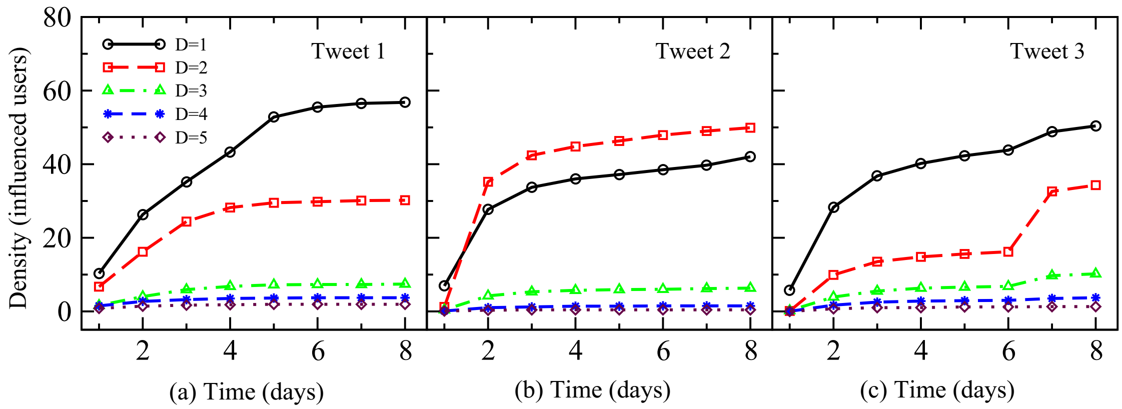

Fig. 3. (Color online) Density of the influenced users for three type of tweets over 8 days with friendship hops as distance.

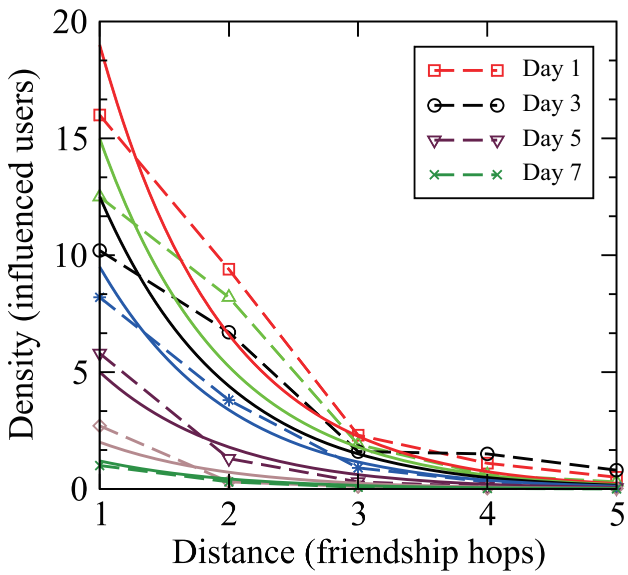

Fig. 4. (Color online) Model results vs. Actual data of the density of influenced users for tweet1 with distance for different time from 1 to 7 days.

to 0. We also choose 50 time steps of the hydrodynamic model correspond to 1 days of the information spreading, this choice takes into account both result accuracy and the model efficiency. The 100 space steps of the model correspond to the initial radius of the source and 1 distance of friendship hops.

To quantify the accuracy of the model results, we calculate the accuracy, A_p, of the hydrodynamic model as follows,

$$A_p = \frac{|V_p - V_a|}{V_a}, \tag{3}$$

where V_p is the model value and V_a is the actual data value. In table III we list the accuracy for model results by comparing with the actual data of distance 1 to 5 and time of days 1 to 6 of the tweet 1. The hydro-IDP model gives the average accuracy of 81.82 percent at distance 1 and 76.70 percent for the first three hops.

Our studies have shown that by using hydro-IDP model with a few proper parameters, one can model the information diffusion and represent the active data accurately. The model parameters could offer a reference for predicting the information diffusion. If we can classify the information population, we can further use the hydrodynamic model with the parameters of user influence and the social platform diffusivity to predict the information diffusion.

IV. Conclusions and Future Work

In this article, we have proposed a hydro-IDP model to represent the information diffusion in the mobile social networks. Through measuring the density of influenced users at the certain distance and the certain time, we characterize the information spreading process in both temporal and spatial dimensions for a real dataset collected from Sina micro-blog website. And then we model the information diffusion by using the hydrodynamics and study carefully the contribution for promoting information diffusion by the information popularity, user influence and the diffusivity of the social platform. Our results show that the model achieves a prediction accuracy of 76.70 percent for the majority of tweets in the real dataset. This result indicates that one can further predict the influence of a given information with a few of normal parameters which could be determined with the past dataset.

In the next step, we plan to extend the hydro-IDP model by considering the superimposed effect for several influential users in the information spreading process. The extended hydro-IDP model also hopefully model the cross-platform diffusion of the information. In addition, the flow temperature corresponds the information popularity at a certain time and distance, and the information diffusion rate with the popularity is also of interests.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (grant No. 61572220).

References

[1] X. Ge, S. Tu, G. Mao, C.-X. Wang and T. Han, "5G Ultra-Dense Cellular Networks," *IEEE Wireless Communications*, Vol. 23, No. 1, pp.72-79, Feb. 2016.

TABLE II
THE ACCURACY OF THE PREDICTION WITH HYDRO-IDP MODEL.

Distance	t=1	t=2	t=3	t=4	t=5	t=6	Average
D=1	81.33	86.56	77.58	84.22	86.27	74.93	81.82
D=2	70.31	64.63	65.71	89.63	62.47	58.66	68.57
D=3	98.03	94.19	93.72	71.94	50.18	52.14	76.70
D=4	67.82	98.48	67.33	75.24	68.72	90.10	77.95
D=5	45.66	61.94	50.57	78.14	67.82	65.94	61.51

[2] X. Ge, H. Cheng, M. Guizani, T. Han, “5G Wireless Backhaul Networks: Challenges and Research Advances,” *IEEE Network*, Vol. 28, No. 6, pp. 6-11, Nov. 2014.

[3] A. Guille, H. Hacid, C. Favre and D. Zighed, “Information Diffusion in Online Social Networks: a Survey”, *SIGMOD Record*, Vol. 42 pp.17-28, 2013.

[4] M. Rodriguez, L. Song, N. Daneshmand, and B. Schölkpof, “Estimating Diffusion Network Structures: Recovery Conditions, Sample Complexity and Soft-thresholding Algorithm”, *International Conference on Machine Learning (ICML)*, 2014.

[5] J. Yang and J. Leskovec, “Modeling information diffusion in implicit networks”, *In ICDM*, Vol. 10, pp. 599-608, 2010.

[6] J. Yang, and S. Counts, “Comparing Information Diffusion Structure in Weblogs and Microblogs”, *4th Intl AAAI Conference on Weblogs and Social Media*, 2010.

[7] S. Myers, C. Zhu and J. Leskovec, “Information Diffusion and External Influence in Networks”, *KDD 12 Proceedings of the 18th ACM, SIGKDD international conference on Knowledge discovery and data mining* pp. 33-41.

[8] Chunxiao Jiang, Yan Chen, Liu K.J.R., “Modeling information diffusion dynamics over social”, *IEEE International Conference on networks, Acoustics, Speech and Signal Processing (ICASSP)*, 2014.

[9] C. F. Lai, H. C. Chao, Y. X. Lai, and J. Wan, “Cloud-assisted realtime transrating for http live streaming,” *IEEE Wireless Communications*, vol. 20, no. 3, pp. 62-70, 2013.

[10] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, “Characterizing User Behavior in Online Social Networks”, in Proceedings of *ACM SIGCOMM International Measurement Conference*, November 2009.

[11] C. F. Lai, H. Wang, H. C. Chao, and G. Nan, “A network and device aware qos approach for cloud-based mobile streaming,” *IEEE Transactions on Multimedia*, vol. 15, no. 4, pp. 747-757, 2013.

[12] C. F. Lai, R.-H. Hwang, H.-C. Chao, M. Hassan, and A. Alamri, “A buffer-aware http live streaming approach for sdn-enabled 5g wireless networks,” *IEEE Network*, vol. 29, no. 1, pp. 49C55, 2015.

[13] G. Ver Steeg, R. Ghosh and K. Lerman. ”What Stops Social Epidemics?”, Proceedings of *the Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[14] N. Du, Y. Liang, N. Balcan and L. Song, “Influence Function Learning in Information Diffusion Networks”, International Conference on *Machine Learning (ICML)*, 2014.

[15] S. Tang, N. Blenn, “Christian Doerr and Piet Van Mieghem, Digging in the Digg Social News Website”, *IEEE Transaction on Multimediam*, 2011.

[16] Sadikov E, Medina M, Leskovec J, et al. ”Correcting for missing data in information cascades”, in Proceedings of *the fourth ACM international conference on Web search and data mining*. ACM, pp. 55-64, 2011.

[17] J. langa, J. Robinson, A. Rodriguez-Bernal and A. Suarez, “Permanence and asymptotically stable complete trajectories for nonautonomous lotkavolterra models with diffusion”, *SIAM J. MATH. ANAL.* Vol. 40, pp. 2179-2216, 2009.

[18] K. Lerman and R. Ghosh, “Information Contagion: an Empirical Study of Spread of News on Digg and Twitter Social Networks”, in proceedings of *4th International Conference on Weblogs and Social Media (ICWSM)*, 2010.

[19] M. D. Choudhury, Y. R. Lin, H. Sundaram, et al. ”How does the data sampling strategy impact the discovery of information diffusion in social media?”, *ICWSM*, Vol. 10, pp. 34-41, 2010.

[20] R. Ghosh and K. Lerman, “A framework for quantitative analysis of cascades on networks”, *ACM International Conference on Web search and data mining*, 2011.

[21] M. Farajtabar, M. Rodriguezy, Nan Du, et al. ”Back to the Past: Source Identification in Diffusion Networks from Partially Observed Cascades”, *Artificial Intelligence and Statistics (AISTATS)*, 2015.

[22] F. Wang, H. Wang and K. Xu, “Diffusive logistic model towards predicting information diffusion in online social networks”, *32nd International Conference on Distributed Computing Systems Workshops (ICDCSW)*, pp. 133C139, 2012.

[23] F. Wang, H. Wang, K. Xu, J. Wu and J. Xia, “Characterizing Information Diffusion in Online Social Networks with Linear Diffusive Model”, *33nd International Conference on Distributed Computing Systems (ICDCS)*, pp. 307-316, 2013.

[24] W. Galuba, K. Aberer, D. Chakraborty, et al. ”Outtweeting the twitterers-predicting information cascades in microblogs”, in proceedings of *the 3rd conference on Online social networks*, USENIX Association, pp. 3-11, 2010.

[25] A. Nematzadeh, E. Ferrara, A. Flammini, and Yong-Yeol Ahn, “Optimal Network Modularity for Information Diffusion”, *Phys. Rev. Lett.*, Vol. 113, 259901, 2014.

[26] A. Guille, H. Hacid, “A predictive model for the temporal dynamics of information diffusion in online social networks”, in proceedings of *the 21st international conference companion on World Wide Web*. ACM, pp. 1145-1152, 2012.

[27] M. G. Rodriguez, J. Leskovec and A. Krause, “Inferring networks of diffusion and influence”, in proceedings of *the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 1019-1028, 2010.

[28] F. Jin, E. Dougherty, P. Saraf, Y. Cao, N, Ramakrishnan, “Epidemiological Modeling of News and Rumors on Twitter”, in proceedings of *the 7th Workshop on Social Network Mining and Analysis*, Article No. 8, pp. 551-556, 2013.

[29] H. Wang, F. Wang and K. Xu, “Modeling Information Diffusion in Online Social Networks with Partial Differential Equations”, arXiv:1310.0505.

[30] L. D. Landau and M. Lifshitz, “Fluid Mechanics”, Pergamon press.

[31] J.D. Murray, “Mathematical Biology I. An Introduction”, *Springer-Verlag*, New York, 1989.

[32] M. Holt, “Numerical methods in fluid dynamics”, *Springer Series in Comput. Physics* (Springer, Berlin, 1977).

[33] V. Schneider, U. Katscher and D. H. Rischke. ”New Algorithms for Ultra-relativistic Numerical Hydrodynamics”, *J. Comp. Phys.*, Vol. 105, pp. 92-107, 2003.