DOI: 10.15503/emet2015.71.78

Natural vs artificial viral spread within the online community

Andrzej Jarynowski
Department of Theory of Complex Systems, Smoluchowski Institute,
Jagiellonian University, Łojasiewicza Street 11, Kraków, Poland
Department of Theoretical Physics, Moldova State University,
Alexei Mateevici Street 60, Kishinev, Moldova
Jarosław Jankowski
Department of Computer Science and Information Technology, West
Pomeranian University of Technology, Żołnierska Street 49, Szczecin, Poland
Anita Zbieg
University of Wroclaw, Institute of Psychology,
Dawida 1. Wroclaw, Poland
E-mail address: andrzej.jarynowski@uj.edu.pl

Abstract

Diffusion of information and viral content, social contagion and influence are still topics of broad evaluation. We have studied the information epidemic in a social networking platform in order to compare different campaign setups. The goal of this work is to present new knowledge obtained from studying two artificial (experimental) and one natural (where people act emotionally) viral spread that took place in a closed virtual world. We propose an approach to understand the behavior of an online community exposed to external impulses as an epidemic process. The presented results are based on online multilayer system observation, and show characteristic differences between setups. Moreover, some important aspects of branching processes are presented. We ran experiments, where we introduced a virus to the system and agents were able to propagate it. There were two modes of experiment: with or without award. The dynamic of spreading both of viruses were described by epidemiological model and diffusion. Results of experiments were compared with a real propagation process - spontaneous organization against ACTA. During general-national protest against new anti-piracy multinational agreement – ACTA (criticized for its adverse effect on e.g. freedom of expression and privacy of communication), members of the chosen community could send a virus such as Stop-ACTA transparent. In this scenario, we are able to capture the behavior of society, when real emotions play a role, and compare results with artificiality-conditioned experiments. Moreover, we could measure the effect of emotions in viral propagation. The theory explains the role of emotions in spreading behavior as a factor of message targeting and individuals spread emotional-oriented content more carefully and in a more influential way. Our experiments suggest that probabilities of secondary infections may be four times bigger if emotions play a role.

Keywords: modeling, viral marketing, social behavior, information diffusion.

Introduction

The use of the Internet has become very popular and it has become an inherent part of users lives (Kuś, 2014). Virtual communities reflect many of the properties of offline social groups and organs. Many experiments have been already conducted in a virtual environment. Via the Internet we could study people's behavior in very cheap, but also in very wide way. People leave incredible amount of information on the Internet (Big Data) and current trends in sociology is to analyze it. A new area of interdisciplinary studies was recently been founded - Computational Social Science (Jarynowski, 2014), mainly because of growing popularity of the Internet. Moreover, we could run experiments in the Internet, which in a standard social environment would be difficult or even impossible to conduct.

In this paper, we focus on epidemiology modeling and we introduce the basic rules of computational methodology using the Internet. The studies that direct attention to disease outbreaks (Anderson, & May, 1992), diffusion of innovation process (Rogers, 2005), social influence mechanism (Nowak, Szamrej, & Latané, 1990), social contagion or cascades of influence patterns (Watts, 2002) usually tackles similar phenomenon: an epidemic spread (e.g. pathogen, information content, opinions, behaviors, emotions) within a network of social relations. Such epidemics may follow a variety of models, depending on the spreading properties, content, incentives, risk, individual attitudes of sender and recipient to the content and other factors (Grabowski, 2012). Due to the fact that the knowledge of such specifics is essential for the epidemiologist or sociologist, researchers constantly evaluate models and real-world data (Liljeros, 2010). Researches in the field of online communities, virtual worlds and massively multiplayer platforms and games allow studying designed experimental setups usually not possible to observe in offline communities (Leskovec, Adamic, & Huberman, 2007). Nevertheless, the empirical network studies seem to be promising, relatively little has been done in this area and the main focus contains random networks (Malarz, Szvetelszky, Szekfu, & Kułakowski, 2006). Recently, for sociology and medicine the knowledge of complex system tools such as networks has undergone an accelerating growth, however all models of such systems are incomplete without real data, especially register-based (Rocha, 2011). The Internet provides opportunities for both gathering and analyzing data. Network theory is useful when it comes to studying nature from a systems perspective, and there are several examples in which it has helped understanding the behavior of complex systems (Liljeros, 2001). Social interactions are a good example in which a network perspective is important to understand systems behavior (Grabowski, 2012). However, viral spread and even society itself change in time at different scales and in different ways. Dynamics from this perspective have not been studied in detail and an integrative framework is missing (Holme, 2012), especially for such an event like Stop-ACTA[7] movement formation. The research presented in

7 The Anti-Counterfeiting Trade Agreement (**ACTA**), a treaty signed in October 2011 to establish international standards for intellectual property rights enforcement.

this paper is targeted at online social platforms with the ability to capture different forms of users' behaviors: communications, activity and goods transfers among users (Czaplicka, & Hołyst, 2012). The main motivation in the current research is to observe human action systems more by analyzing their behavior related directly to viral campaign and stages of the participation for the viral action

Experimental setup and background framework

During the research, there were data from two artificial (sending an avatar gesture) and one natural (sending a Stop-ACTA transparent or mask) viral actions with different characteristics from social platforms working in a form of virtual world situated in one of the Polish Internet virtual worlds. In this virtual world, users can wear avatar [Pic. 1 left] and those items are the core of our study. In spreading interest users can get access to specific elements of the avatar only by invitation, which can be spread without any limitations to other active users. In all actions, users were spreading virtual goods using the viral mechanism to their friends. The first viral action was based on sending gifts to friends and the senders' motivation to spread those gifts was not incentivized. The second was based on incentives and competition among users to spread visual elements of avatars among their friends. The best spreaders would receive prizes as well as some randomly chosen invitations in gratitude. The third – natural study was performed when Stop-ACTA movement[8] was starting in Poland. Stop-ACTA stand up [Pic. 1 right], which waved emotional value, could be spread.

Pic. 1. Visualization of different avatars (left) and Stop-ACTA stand up (right)
Source: Own research.

8 After Poland's announcement that it will ratified the ACTA treaty, protests were held against, mainly conducted by Internet – users like our studied community

Study design

Epidemic spread can take place between individuals who form different kinds of social networks. Empirical research related to the question of factors driving social contagion come from diverse studies, but mostly based on networks. Nowadays, based on data gathered in computer systems, a new type of social network can be extracted and analyzed (Liljeros, 2012). These networks are usually automatically extracted from such data sources as: bibliographic data, blogs, e-mail systems, telecommunication data, and social services like Twitter or Facebook (Goel, Watts, & Goldstein, 2012), video sharing systems like YouTube (Liu-Thompkins, 2011), Wikipedia and many more.

In a current study, an individual become infected only by another individual. Thus, the spread of a virus can be described in terms of the epidemiological process. Thus, let's use epidemiological notation of SEI (susceptible – exposed – infective) concept (Anderson & May, 1992). In this study, sending invitation is like transmitting pathogen from I to S and the receiver changes his state from S to E. If the agent, who was in state E, decides to send further invitation, he becomes to be in state I. The process is irreversible. Because infective agents could send invitations to any active users (in both possible states S, E, I), we can differentiate two kinds of infections: unique (when agent in state S get his first invitation) and non-unique.

$$\boxed{S \xrightarrow{pe} E \xrightarrow{p} I} \qquad \boxed{S \xrightarrow{pe} E} \qquad \boxed{S \xrightarrow{pc} I}$$

Pic. 2. Model definitions: full model (left), reduced model where transmission from E to I omitted (center), reduced model where state E is skipped (right)
Source: Own research.

Concept of data analysis

In a virtual world, exchange of users (periodic day-night patterns) makes observations more complicated (Grabowski & Kosinski, 2010). To cope with that, let us investigate the process only on a global scale. To get a general view and find general properties we observe the whole campaign from a stationary, and not time dependent perspective. We focus on global parameters of an epidemic. Although, we know they are not static and are changing with time (Jacob, 2010) (waves or generations) [Fig. 1], but only this level comparison between campaigns gives us an opportunity to conclude something about the difference in a quantitative manner.

Additionally, to get more insight in a dynamic time framework we observe the first day of campaign [Fig. 2]. This approach gives opportunity to a reader to understand what is happening in the system (branching process) in more qualitative way.

In the current study, to expose the role of emotions in viral spread we choose to consider only small network snapshots, but in other papers about this virtual world (Jankowski, 2011) you can find other network aspects (Jankowski, Ciuberek, Zbieg, & Michalski, 2012), (Zbieg, Żak, Jankowski, Michalski, & Ciuberek, 2012) used additionally for predicting effects of marketing strategies (Jankowski, Michalski, & Kazienko, 2012), (Michalski, Jankowski, & Kazienko, 2012).

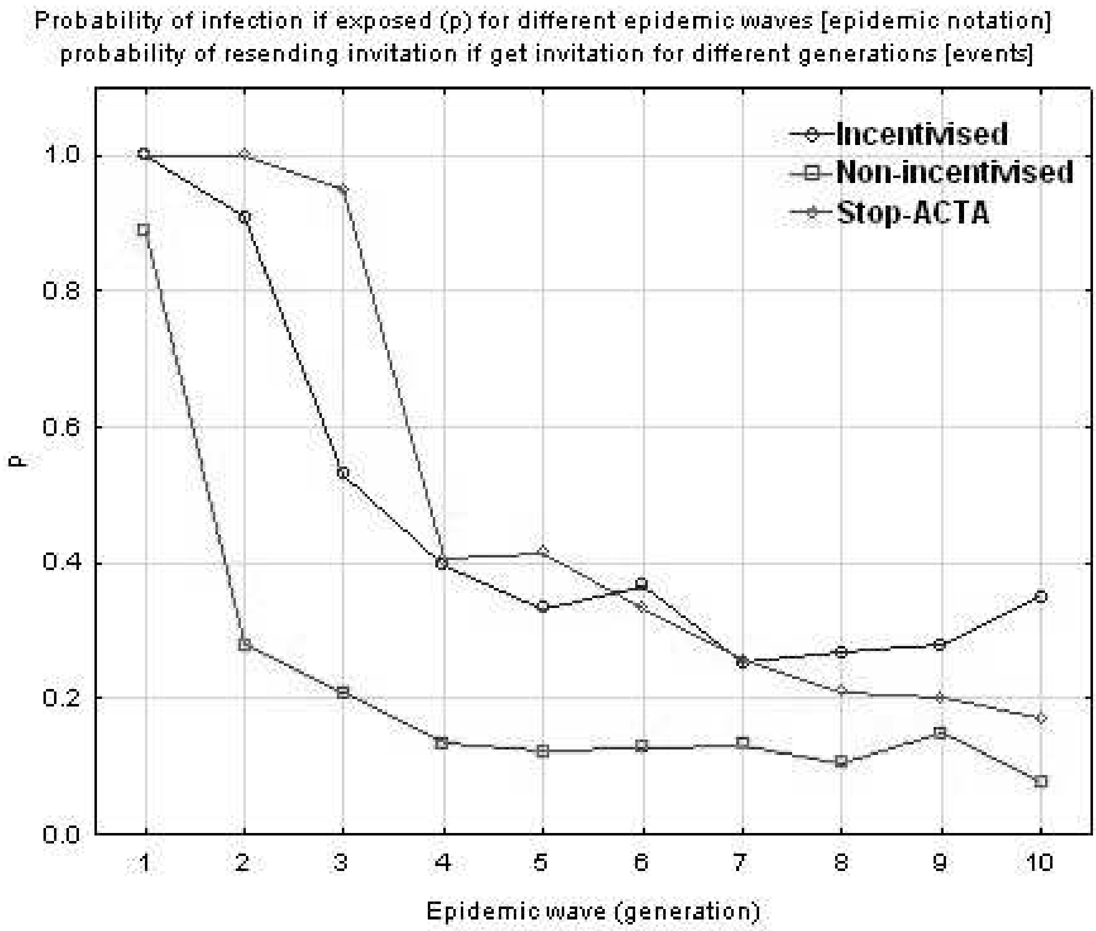

Fig. 1. Dynamics of campaign property - probability of infection if exposed (*p*) / probability of resending invitation if get invitation - for first 10 generations.
Source: Own research.

Results

Overall statistics of campaigns [Tab. 1] seem to differ in an ordered way. Unfortunately, we have only one realization of each scenario, so the random effect would play a big role in terms of measured variables. More informative are model parameters [Pic. 2, Tab. 2], which explain ruling behavior. The main difference for emotional study (Stop-ACTA) is observed for *pe* and *pc* values. Those probabilities are a few times bigger for natural experiment.. Using epidemiological notation, probability to send infection for a pair of individuals who were in contact (*pe*) is much bigger for Stop-ACTA study. It can be understood, that viruses were sent in a more precise way. Additionally, big value of probability of infection per contact (*pc*) means that such transmissions (invitations) are effective. Those two parameters tell us that emotional oriented information is targeted. In other words, we can claim that people are sending emotional information mainly to people, for whom that message can be relevant. On the other hand, we have to add that probabilities of infection if exposed (*p*) are similar (difference up to 50%), so probability of sending infection further does not depend as much on emotional aspect. Standard deviation calculated based on aggregated generation data [Fig. 1] does not allow claiming any conclusions with statistical significance. As an example, we also can suggest, that probabilities of retwitting (in Twitter or Facebook) would not depend on the emotional value of information (how it seems to be observed (Thelwall, Buckley, & Paltoglou, 2012).

Table 1. Campaign quantities

Event	all invitations	unique invitations	No. users who send invitation	No. of generations (until no more is send)	Time of campaign
Epidemiological notation	Total. Contacts	Total . Exposed (E)	Total Infective (I)	No. waves (until epidemic died out)	Time of epidemic
Non-incentivized	9972	3069	746	12	13 days
Stop-ACTA	**731**	**635**	**242**	**14**	**44(5) days**
Incentivized	28446	3874	1873	14	13 days

Source: Own research.

Table 2. Model parameters calculated for campaigns

Event	proportion of unique invitations	probability of resending invitation per invitation	probability of resending invitation if get invitation
Epidemiological notation	probability of being exposed per contact (*pe*)	probability of infection per contact (*pc*)	probability of infection if exposed (*p*)
Non-incentivized	0.31	0.07	0.24 (std=0.23)
Stop-ACTA	0.87	0.33	0.38 (std=0.28)
Incentivized	0.14	0.07	0.48 (std=0.26)

Source: Source: Own research.

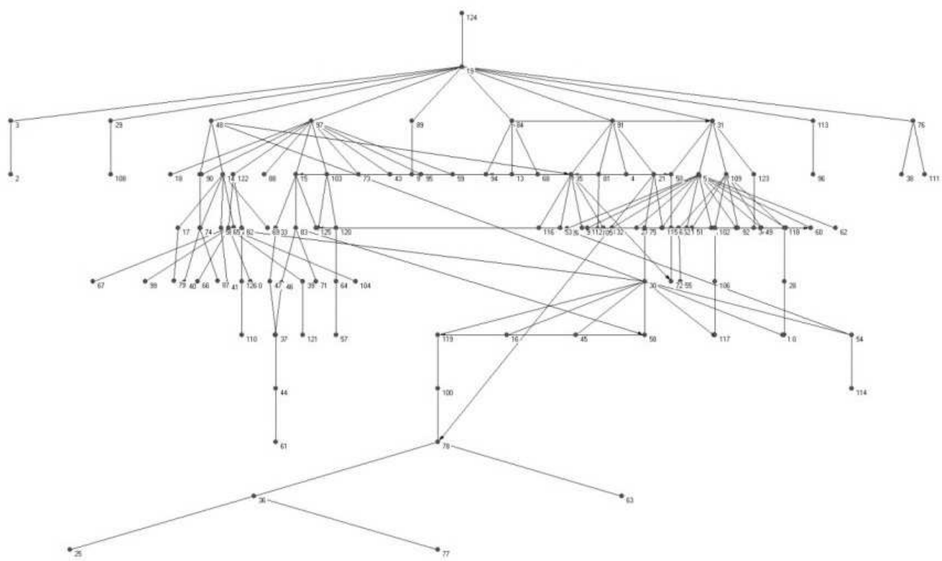

Fig. 2. Hierarchical representation of Stop-ACTA campaign during the first day. Links show branching process and arrows non-unique infections (invitations).

Source: Own research.

We present graph [Fig. 2] of 53 nodes – users (who get exposed - E) who sent further infection as a tree (Burda, Correia, & Krzywicki, 2001). During first day of Stop-ACTA campaign 10 waves (generation) come to life before process dies out (only for a night, because the virus was propagating since the morning the day after but local analysis stopped only on the first day).

Conclusions and further work

The presented research showed a epidemiological approach (Jarynowski, 2010) to viral campaign analysis by comparing emotional with artificial experiments. We show possibilities of computational methodologies within experimental research in the Internet. Social contagion studies are usually a simplification of a real world situation. Decisions related to participation in viral diffusion are based on many different factors that are difficult to observe and monitor. Even then we observed that emotional messages [Tab. 2] are deliberately targeted (high *pe* and *p*c for Stop-ACTA) and less blindly forwarded (the same *p* for Stop-ACTA and both artificial). On the other hand, our conclusions are still below any statistical significance [Tab. 2] and external factors like media, mainly on Stop-ACTA campaign, cannot be controlled. By global analyses, it is possible to catch some deeper characteristics and interesting results. The study showed that model parameters are time or generation dependent [Fig. 1]. Ongoing research will be focused on a network-oriented agent based model of viral spread. Message exchange between users A and B, entries to individual profile, etc. will able us to constrain contact network, which was unknown in the presented model. Finally, the properties of individuals, like frequency of expressing negative of positive emotion can be measured (e.g. by counting emoticons) and applied to model.

References

Anderson, R. M., & May, R. M. (1992). *Infectious Diseases of Humans: Dynamics and Control.* Oxford: Oxford University Press.

Burda, Z., Correia, J., & Krzywicki, A. (2001). Statistical ensemble of scale-free random graphs. *Physical Review E*, *64*, 461181-461189.

Czaplicka, A., & Hołyst, J. (2012). Modeling of Internet Influence on Group Emotion. *International Journal of Modern Physics C*, *23*, 1250020.

Goel, S., Watts, D. J., & Goldstein, D. G. (2012). The Structure of Online Diffusion Networks. *ACM Journal*, *February*.

Grabowski, A. (2012). The relationship between human behavior and the process of epidemic spreading in a real social network. *EPJ B*, *85*, 248.

Grabowski, A., & Kosinski, R. (2010). Life span in online communities. *Phys. Rev. E*, *82*, 066108.

Holme, P. (2012). Temporal networks. *Physics Reports*, *519* (3).

Jacob, C. (2010). Branching Processes: Their Role in Epidemiology. *International Journal of Environmental Research and Public Health*, *7*, 1186–1204.

Jankowski, J. (2011). Analysis of Multiplayer Platform Users Activity Based on the Virtual and Real Time Dimension. *LNCS*, *6984*, 312-315.

Jankowski, J., Ciuberek, S., Zbieg, A., & Michalski, R. (2012). Studying Paths of Participation in Viral Diffusion Process. *LNCS* (7710), 503-516.

Jankowski, J., Michalski, R., & Kazienko, P. (2012). The Multidimensional Study of Viral Campaigns as Branching Processes. *LNCS*, *7710*, 462-474.

Jarynowski, A. (2010). Human-human interaction: epidemiology. In *Life time of correlation.* Wrocław: Wydawnicwto Niezależne.

Jarynowski, A. (2014). *Obliczeniowe Nauki Społeczne w Praktyce.* [Computational social science]. Wrocław: WN.

Kuś, J. (2014). Basic principles of conducting psychological experiments. *E-methodology, 1*, 9-18.

Leskovec, J., Adamic, L., & Huberman, B. (2007). The dynamics of viral marketing. *ACM Trans. Web*, *May*.

Liljeros, F. et all. (2012). Implementation of Web-Based Respondent-Driven Sampling among Men who Have Sex with Men in Vietnam. *PLoS One*, *7(11),* e49417.

Liljeros, F. et all. (2010). Information dynamics shape the sexual networks of Internet-mediated prostitution. Proceedings of the National Academy of Sciences, 107(13), 5706-5711.

Liljeros, F. et all. (2001). The web of human sexual contacts. *Nature*, *411* (6840), 907-908.

Liu-Thompkins, Y. (2011). Seeding Viral Content: Lessons from the Diffusion of Online Videos. (unpublished)

Malarz, K., Szvetelszky, Z., Szekfu, B., & Kułakowski, K. (2006). Gossip in random networks. *Acta Physica Polonica B*, *37* (11), 3049-3058.

Michalski, R., Jankowski, J., & Kazienko, P. (2012). Negative Effects of Incentivised Viral Campaigns for Activity in Social Networks, In *Cloud and Green Computing (CGC), 2012 Second International Conference on* (pp. 391-398). IEEE.

Nowak, A., Szamrej, J., & Latané, B. (1990). From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review*, *97*, 362-376.

Rocha, L. (2011). *Exporing patterns of empirical networks.* PhD Thesis, Umea University.

Rogers, E. (2005). *Diffusion of Innovations.* New York: Free Press.

Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social Web. *Journal of the American Society for Information Science and Technology, 63 (1)*, 163-173.

Watts, D. (2002). A simple model of global cascades on random networks. *PNAS*, *99 (9),* 5766-5771.

Zbieg, A., Żak, B., Jankowski, J., Michalski, R., & Ciuberek, S. (2012). In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)* (pp. 1259-1265). IEEE Computer Society.