

Development of Manipulation Detecting Algorithm as a Constituent of the Internet Community Management System

Zoriana Holub

Social Communication and Information Science Department,
Lviv Polytechnic National University, UKRAINE, Lviv,
S. Bandery street 12, E-mail: zorianaholub@gmail.com

The aim of the research is to outline main peculiarities of the manipulation in the Internet communities. The research encompasses online communication platforms which support an interactive text communication. In the paper feasibility of exploiting traditional schemes of protection from manipulation in the virtual space is questioned, differences of manipulation in real life and in the Internet communities are made obvious. On the basis of the research the algorithm for detecting manipulation in the Internet communities is suggested. The later uses technical features, peculiarities of accounts representations and content of messages to identify the fact of manipulation in the Internet community.

Key words – manipulation, Internet community, IP-address, transaction, frame, neurolinguistics programming.

I. Introduction

Openness, accessibility as well as ease of use of the Internet are the main reasons of virtual communities popularity as means of information transferring. Discussions themes of virtual communities encompass virtually all spheres of human interest, in particular politics, technical support, health, pastime. A number of people treat virtual communities as sources of news and advice. Moreover, they often appear to be an aid to psychological need of socialization. Social researches show a great number of people to regard information in virtual communities as ultimately correct, objective and exhaustive.

Despite the anonymous nature of messages in virtual communities, they have greater impact on the decision making process of virtual community members than their friends, relatives and colleges. As a result, virtual communities are used with acquisitive purposes. Hidden influence often aims at achieving either marketing and political goals, or influencing financial and business decisions making process.

Described above harmful influence is not the only unfavorable effect on the virtual community caused by manipulation. Online dishibition effect is a tendency of individuals to use more offending and aggressive language than they do in real communication. Therefore it is easier to provoke people to inadequate communication in the virtual space than in real. Such outrageous behavior could have destructive effect on the community itself, as it destroys the positive image of the community, which means decrease in its active user and guests.

Manipulation is not a newly emerged phenomenon. Plenty of scientists devoted their works to manipulative situations, tactics, strategies and tools. They elaborated methods and algorithms of manipulation prevention and

counteraction. Technical features of virtual communities brought about differences in virtual and traditional communication. Resulting from changes in communicational environment, evolution of manipulation took place. Schemes and algorithms of protection from real world manipulation cannot be applied to resist manipulative strategies and tools used in virtual communities.

From the multitude of miscellaneous types of Internet communication platforms, social networks, blogs and forums are the most preferred. The information in these virtual communities is conveyed mainly by means of text messages. Thus, it is reasonable to start the elaborating of manipulation preventing system from developing algorithms for detecting manipulation in text messages of virtual communities.

II. Brief Review of Related Researches

A multitude of Ukrainian scientists investigated the phenomenon of manipulation in the Internet. Ivanytska, Klotschko studied the Internet functional as a platform for manipulation [15], Lisovskyi looked into the problem of the Internet as a means of manipulation in Ukrainian context [16]. Savonchak devoted his works to manipulation methods and tactics.[17]

III. Research Aim

The research is aimed at investigating the phenomenon of manipulation on forums, blogs and social networks, in particular to find similarities and outline difficulties of manipulation in real and virtual communication. Special attention is paid to characteristic features of Internet manipulation, which could be further exploit to develop an algorithm of manipulation prevention and detection.

IV. Peculiarities of Manipulation in the Internet Communities

Various types of manipulation occur in the Internet. They can be categorized into audial (speech, music, sounds), visual (text, graphical items) and combined (video, multimedia). Other classifications are into active or passive, gradual or instant, long-lasting and short-lasting. The research is dedicated to interactive text manipulations in the Internet. To put it differently, under research are manipulation acts performed in the Internet by means of short text messages exchange between a few agents. These messages are usually organized into logical sequences and are subjective. Interactive text messages manipulations often take place on forums discussions, in comments on news web-sites, blogs, company and reference web-sites or online shops.

The feature of interactivity is by stressed as a key in the investigated type of manipulation. Manipulation methods in the messages are not designed for passive perception, they obtain triggers, program the reaction of recipients and exploit it for achieving certain manipulative goals.

In interactive text manipulations in addition to verbal means could be use following types of nonverbal:

- metagraphemics (emoticons, calligrams and hypergraphics); [5]

- links (although they constitute of character, but they have no semantic meaning);
- image (is expressed by an avatar, status or other similar profile details that are bound along with all messages of the author).

The most widely-used means of manipulation in an interactive text communication are linguistic traps. Linguistic traps are implicit constraints of the meaning, realized by use of selective approach to words, rhetoric tricks, symbolic exaggerations and ritual. [6] Among the scientists who investigated linguistic means of manipulation are Teun A. van Dijk, Bryant, Thompson, and Blakar. [12, 13]

In case of interactive text manipulations nonverbal means are complimentary tools, their exclusive usage is not sufficient for achieving proper manipulative results.

V. The object of manipulation in the Internet communities

Due to the fact that information of the Internet communities discussions is easy of access, such type of manipulation as interpersonal manipulation does not occur in the Internet communities. It is possible to assume that a manipulator conducts a preliminary analysis of community participants aiming at determining an object of a future manipulation, however it is efforts-consuming and that is why rarely occurs.

In the Internet manipulations are exploit exclusively universal vulnerabilities (buttons), which exist virtually in every human-being. This ensues for the mass nature of online communication. Especially, at the outset of manipulation there is no exact object of manipulation, a manipulator appeals to the prototype of an object. In other words to the certain group of people possessing suitable for manipulation qualities.

When “the right button” is pressed and the object reacts as foreseen, the object of manipulation becomes less abstract. Instead of targeting a prototype a manipulator applies the strategy to the most yielding members of the Internet community. However, a manipulator keeps an eye on other members as their activities in a community could destroy all manipulative plans.

Another characteristic feature of the Internet community is the absence of credible information on social position, career level, age etc. These and similar feature are often exploit in traditional manipulations. Despite the fact, that there is some kind of hierarchy in social communities (e.g. novice, regular) this differentiation is not equal to the above enumerated status features of a traditional manipulation.

VI. Internet-Communities Members As Manipulation Means

There two reasons of ineffectiveness of the manipulation carried out by use of one agent. First of all, excessive activity in an online community is suspicious. Furthermore, the band wagon is a very effective manipulation technic.

Subtle and multistage manipulations require several agents. The necessary amount of agents is achieved in the following ways: building a group of manipulators or provoking needed actions of community members. In terms of manipulation network structure manipulations are classified into the Internet communities:

- Direct
 - one-agent;
 - group;
- Indirect
 - chain;
 - virus.

Direct manipulations are done by specialists in pursue of a set goal. Depending on the number of agents they are subdivided into one-agent and group manipulations.

Direct one-agent manipulation is carried out from one account. This type of manipulations is used only for fulfilling easy tasks. It is the most easy to detect manipulation. Its popularity is explained by little expense needed for its implementation.

Group manipulation is performed either by several manipulators each in charge of one account or by one manipulator controlling several accounts.

Indirect manipulation is more complicated to commit as well as to detect. To accomplish it successfully a branching scenarios should be elaborated. Manipulation in operation should be monitored on a regular-basis in order to detect deviation and take appropriate actions. The scale of effect of indirect manipulation is much larger than of direct one-agent manipulation.

Indirect manipulation means getting somebody to do your dirty work. In case of indirect manipulation distorted reality is created by the messages of community members who do not belong to manipulators team. A manipulator provokes a user to post messages with appropriate content. Thus, in indirect manipulation triggers that should accomplish the goal of manipulation are embedded in messages of community users. In other words manipulator imposes his idea and encourages community members to disseminate it.

To sum up, when detecting manipulation, it is necessary not only to detect a message with manipulative content and ban the author, but to check whether author of a message is a manipulator or a compromised member of a community. Otherwise the result of such combatting manipulation would be not communities without manipulators, but communities without members.

VII. The algorithm for detecting manipulation

The algorithm for detecting manipulation in the Internet communities consists of two stages. At these two stages discussion is analyzed from different angles and different technics are applied. The two main stages of the algorithm of detecting manipulation are:

- detecting manipulator at the stage of registration;
- detecting manipulation in operation.

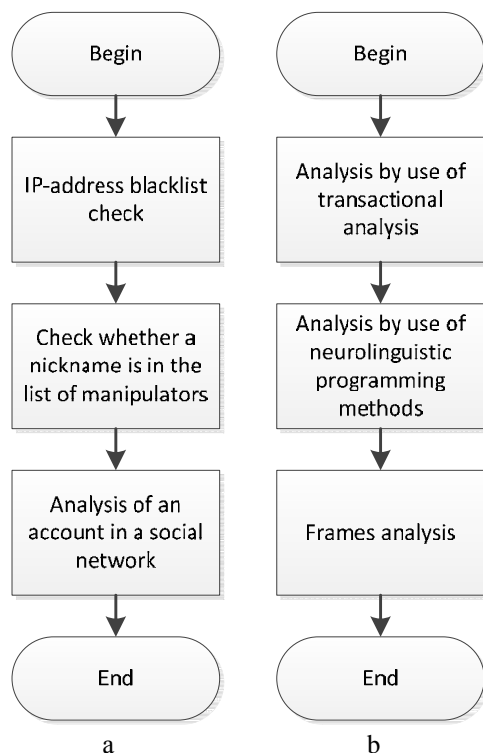


Fig. 1. Detecting manipulators:
 a) by means of account features;
 b) semantic analysis of message's content.

At the first stage technical features of the prospective account (IP-address) and its peculiarities (nickname, profile in social networks) are checked.

The IP-addresses check means detecting previously identified and banned manipulators' accounts. To conduct the check the data base of manipulations in the Internet communities is required. This data base should be updated as new manipulation cases detected.

Furthermore, it is reasonable to check nicknames of community members. This check lies in comparing nicknames of community members with manipulators' nicknames contained in the database. By means of this check it is possible to ban manipulators networks.

Some Internet communities give opportunity to log in with Facebook account. In this case one more method of detecting manipulators is possible: to analyze the prospective member's profile in social networks. That means to use the link and go to the profile and assess the development level of the account. A little personal information, friends, photos are characteristics of the fake account used with a certain manipulative aim.

The second stage of the algorithm for detecting manipulation in the Internet communities is the analysis of linguistic features of a message. This analysis is more complicated than technical check and accounts peculiarity analysis of the previous stage. At this stage are applied various physiological methods, in particular transactional analysis, neurolinguistics programming etc.

At this stage the sequence of messages of a certain member is analyzed for containing any of seven propaganda devices distinguished by Lasswell, namely the name calling, the glittering generalities, the transfer,

the testimonial, the plain folks, the card stacking, the band wagon.[11] Applying quantitative content analysis of the Internet communities discussion verbal and nonverbal means used for constructing each device are defined. For example, when, expressing opinion, a member uses virtue words, such as God, truth, freedom, honor, human rights, than this member is suspected in using the glittering generalities with the aim to accomplish a certain manipulative goal.[10]

One more widely-used manipulation technique ia known as "red herring" (in Lasswell's classification it is called cards stacking).[10] A red herring is something that misleads or distracts from a relevant or important issue. [14] Manipulators use underemphasis and overemphasis to dodge issues and evade facts.

Another approach to detecting manipulation is by means of transactional analysis. The whole discussion is analyzed in light of transactional analysis. The used types of transactions (complimentary, crossed, angular, duplex) are analyzed. Complimentary transactions are characteristic for smooth communication without blaming, flame or invectives. Types of transaction used to fulfill a manipulation are angular and duplex. Both these types a apart from a vector on social (ostensible) level, they have on psychological one. To put it differently these ulterior transactions involve the activity of more than two ego states simultaneously. In angular transaction at social level a transaction is directed to one ego state, while psychological vector is directed by the well-trained and experienced manipulator to another ego state. The manipulator, though, sends both social and psychological stimuli from the same ego state. In duplex transactions not only the ego states of receiver are different on the psychological and social levels, but manipulator also sends stimuli on social and psychological levels from different ego states. [4,8]

Neurolinguistic programming describes various tactics of intentionally hidden influence on others attitudes and decisions. Among these methods is tuning, anchoring, modeling. Tuning, a popular NLP technique is the process of replicate behavior, language, etc. of objects of manipulation in order to convey an ulterior message to their subconsciousness. [1,2] According to neurolinguistics programming people have an internal preferred representational system and preferred to process information primarily in one sensory modality. This sensory modality could be ascertained from external cues (eye movements, posture, breathing, voice). These external cues of sensory modality are not obvious in the Internet communication. As a result to state a sensory modality of an object of manipulation verbal sensory cues are used. For example, words "see" and "bright" would be considered visual sensory cues. In contrast "feel", "comfortable" are used most frequently by kinesthetic. In terms of manipulation detecting, when a person is noticed to use visual verbal cues and at same time suddenly shifts to audial, this person is suspected in manipulation. [7] As this sensory modalities shift was supposedly aimed at tuning to another participant of the Internet discussion.

Framing was also adjusted to using as component of an algorithm of manipulation detection. Framing encompasses theoretical perspectives on how individuals and groups, organize, perceive, and communicate about reality. Frames are widely used to manage perceptions of factual information. Considering recipients attitude frames could be positive or negative. According to aims exist Risky Choice Frames, Attribute Frames, Goal Frames. [3,9]

The results of the algorithm should be included into data base. These results would contribute to effectiveness and accuracy of the algorithm. Furthermore, it is possible to define main tendencies of manipulation in the Internet communities.

VIII. Results Verification of Detecting Manipulation in the Internet Communities

Although manipulation in the Internet communities is carried out in the virtual space, it is aimed at obtaining results in the reality. Due to the fact that final results of manipulations are people's actions and events in the real world, it is difficult to assess results of the algorithm.

In Internet-discussions of a community could participate people that never contact in reality. Internet community participants could have miscellaneous geographical location, they could be citizens of different countries. That is way the evidences of carried out manipulations are shifts in common tendencies.

Approximate esteem of the manipulation scale is made on the basis of side-effects and interim results of the manipulation. For instance, influence of a manipulation, which goal was anti-Ukrainian propaganda, could be assessed regarding the number of previously patriotic, but then changed to neutral avatars. Changing of avatars is a side-effect of manipulation. The count of people, who supports certain messages or begin to produce ones with similar content could be used to assess the effectiveness of manipulation. In this case it would be done by use of interim results.

Conclusion

Research results showed detection of manipulation to be a complicated process, thus vital for proper functioning of the Internet communities. As a result of the analysis of characteristic features of manipulation in the Internet communities the algorithm for detecting manipulation is suggested. The two main stages of the algorithm are thoroughly described. Methods applied at the first stage of algorithm are based on technical features and peculiarities of the account. At the second stage to analyze the content of the message on ulterior influence, different approaches are used. At the second stage methodology of transactional analysis, neurolinguistics programing, framing theory and seven common propaganda devices elaborated in the Institute for Propaganda Analysis are exploited.

References

- [1] Bandler R., Grinder J. *The Structure of Magic I: A Book About Language and Therapy*. Palo Alto, CA: Science & Behavior Books, 1975.
- [2] Bandler R., Grinder J. *The Structure of Magic II: A Book About Communication and Change*. Palo Alto, CA: Science and Behavior Books, 1975.
- [3] Bandura A. *Social Cognitive Theory of Mass Communication*. *Media Psychology* 3, pp.265-299, 2001.
- [4] Berne E. *Games People Play – The Basic Handbook of Transactional Analysis*. New York: Ballantine Books. 1964.
- [5] Blakar. R.M. *Language as a means of social power*, In: *Pragmalinguistics*, J. Mey (ed.). The Hague—Paris, Mouton, 1979.
- [6] Bryant J., Thompson S. *Fundamentals Of Media Effects*, (First Edition), pp. 194—200, 2001.
- [7] Dilts R., Grinder J., Bandler R., Cameron-Bandle L., Delozier J. *Neuro-Linguistic Programming: Volume I: The Study of the Structure of Subjective Experience*. Scotts Valley, CA: Meta Publications, 1980.
- [8] Goodin R. E. *Manipulatory Politics*. New Haven, Conn.: Yale University Press, 1980.
- [9] Ian Stewart, Vann Joines *TA Today: A New Introduction to Transactional Analysis of Scripts People Live: Transactional Analysis of Life Scripts*. New York: Grove Press By Claude Steiner, 1990.
- [10] Lasswell H. *Propaganda Technique in the World War (1927; Reprinted with a new introduction, 1971)*.
- [11] Lasswell H. *Psychopathology and Politics*, (1930; reprinted, 1986).
- [12] Rudinow J. *Manipulation, Ethics*, vol. 88, no. 4, pp. 338-347, 1978.
- [13] Teun A. van Dijk *Pragmatics of language and literature*. Amsterdam: North Holland, 1976.
- [14] *Oxford English Dictionary*. red herring, n. Third edition, September 2009; online version December 2011. <http://www.oed.com/view/Entry/160314>; accessed 21 Щлещиук 2015. An entry for this word was first included in *New English Dictionary*, 1904.
- [15] Клочко О.В., Іваницька О.О. *Сучасні інтернет-технології як засіб маніпуляції свідомістю в сучасному суспільстві*. XXIV Київський міжнародний симпозіум з наукознавства та науково-технологічного прогнозування, м. Київ, Україна, 2-3 червня 2011р.
- [16] Лісовський П. М., *Маніпуляція свідомістю: сутність, структура, механізм у сучасному трансформаційному суспільстві (соціально-філософський аналіз) : Моногр. К. : Вид-во Нац. пед. ун-ту ім. М.П.Драгоманова, 200, с2006*.
- [17] Савончак, В. Я. *Особливості структури журналістського сайту: на прикладі порталу risu.org.ua С. 73-77*