

Vulnerability Profile of Online Social Network Users

Mikhail Yu. Monakhov

Dept. of Informatics and Information Security
Vladimir State University
Vladimir, Russia
mmonakhov@vlsu.ru

Yuri M. Monakhov

Dept. of Informatics and Information Security
Vladimir State University
Vladimir, Russia
unklefc@gmail.com

Andrey V. Telny

Dept. of Informatics and Information Security
Vladimir State University
Vladimir, Russia
andre.izi@mail.ru

Abstract — This article presents an approach to the vulnerability profiling for online social network users. The profile includes information about the socio-demographic group of the user, calculated indicators of the user's emotionality by eight scales: "Inspiration", "Uncertainty", "Anger", "Discontent", "Joy", "Fear", "Indifference", "Disorder". The authors propose a method for estimation of emotional indicators based on the users' message analysis. Associative experiment, the method of thematic classification, semantic differentials, ontology have been used in the analysis of texts. We use the PMI (Pointwise Mutual Information) index as a numerical characteristic of emotionality. The authors analyze the pilot research results and make recommendations concerning the application of the proposed approach.

Keywords — social network, information threat, emotional scaling, semantic differential, vulnerability profile, emotional index, information retrieval.

I. INTRODUCTION

When modeling the processes associated with the implementation of information threats carried out in social networks, i.e. dissemination of misinformation (low-reliability messages, e.g. "fake news"), propagandistic and advertising messages in the course of information warfare, etc., it is necessary to take into account not only the "information force" of messages received by the network user, but also the susceptibility of users to such impacts [1,2].

The threat is implemented in the social environment – therefore, for its successful implementation, the attackers "adapt" to this environment: they know the social portrait of the victim, perhaps one's interests and style of communication.

The task of protecting users is to create their immunity to this type of attack. The social network, as a technical object, can be protected, for example, from messages containing officially prohibited information, through the use of standard content filtering systems. As for the threats of hidden propaganda or misinformation, they can be much more difficult to repel.

Protection becomes individual here. Each user needs to be approached differently, to the extent of knowing his internal properties of susceptibility to informational impacts - his "immune" barrier. If one could identify patterns of construction of such barrier, then this barrier could be successfully controlled.

Below we provide the algorithm of information threat distribution in social networks. Figure 1 depicts the attack scheme.

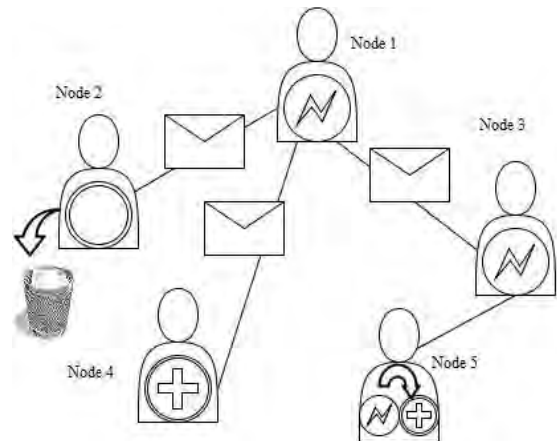


Fig. 1. The attack scheme

Step 1. The attacker initiates information threat distribution (an "attack" process), distributing "dangerous" messages according to his contacts list. Either a single user or a group can perform this attack.

Step 2. Recipients (nodes 2, 3, 4) accept the "dangerous" message further or join the attack process, extending it according to their contact lists (node 3), or ignore or generally delete the message (node 2), i.e. some recipients do not participate in the attack. An attack process can go as avalanche. The attackers do not stop the attack after the one time

transmission of the "dangerous" message. The attack window extends through the quite considerable period and depends on the "dangerous" message type and susceptibility of the user.

Step 3. The attacked users can cease accepting and distributing "dangerous" messages (node 5) due to the influence of information protection mechanisms, for example, the external warning. Therefore, such users will permanently reject these "dangerous" messages.

Step 4. This process continues in OSN while there are attacker nodes, susceptible nodes, and/or security processes are absent.

Therefore, process of information threat distribution in OSN is in essence the complex dynamic process consisting of two counteracting subprocesses: attack and protection of network's nodes.

We will assume that vulnerability is a complex property of a social network user to be susceptible to an information attack. We will assume that "susceptibility" depends on the type of socio-demographic group of the user (age, gender, education, etc.) and the degree of his emotionality –impressionability, sensitivity, impulsiveness. The set of such characteristics for the user or a group of users will be named the vulnerability profile of the user (or the group). The process of information attack consists in an information threat exploiting one or another vulnerability (or their totality) of the user. If the exploitation succeeds, then the user "accepts" a dangerous message, and acts according to an attacker's plans, e.g., passes misinformation further.

II. OVERVIEW

We deem the construction of the vulnerability profile for the online social network (OSN) users technically feasible due to the following key presuppositions and facts:

- viral marketing is a method of disseminating advertisements in a social network, where main distributors of information are the recipients themselves [3]. Information influence is carried out by mass-sending of the message capable to attract information recipients due to emotionality of the text, and uncanniness of the premise. The efficiency of influence increases if there is trust between the recipient and the sender;
- mutual awareness [4]. The OSN user knows about the actions and views of other users, environmental parameters (the state of nature). It influences a set of values, social attitudes and representations available for the user. That influence leads to establishment of the new, or modified old representations. In [5], a coordination game with incomplete knowledge is considered. The authors of [5] allocate minimally sufficient networks, which reinstate users into the hierarchy of social roles: from "initial adopters" and "followers" to the "late adopters". That provides understanding of a role played by general thesaurus in collective action and its connection with the social network structure;

- model of behavior or person's views change under the influence from other people judgments (conformity) in big social group (the Ising model based on thermodynamic analogies [6, 7]). The social network allows users to create virtual users with fictional appearance, character, and history. Intruder can realize threat with high degree of trust working under the guise of such users. The same information can found on several forums, blogs at the same time. Users feels reliability when they reads same message in different source;
- research into the communication processes in OSNs is based on influence models, information management and confrontation [4-12]. Threshold models [9, 10], the independent cascades model, models based on interacting particle systems [11], dynamic Bayesian network with two-level structure – level of individuals and level of group [12], are all used for collective behavioral modeling;
- dynamic of harmful message distribution is described by the SIR (Kermack-McKendrick) models, MSEIR and SEIR, as by analogy to models of epidemics and propagation of computer viruses [13]. In [13], the vulnerability coefficient does not change from user to user and characterizes the OSN topology.

These features represent the foundation for studying messages threats. We will consider the following user's characteristics specified in his profile: gender, age, place of residence, kind of activity and so forth.

However, the level of users' emotional constitution is the definitive factor in estimating the efficiency of influence for information threats. The authors can therefore posit that the problem of estimating the message emotionality as a basic threat feature and estimating of OSN users' emotionality as a basic vulnerability feature, is currently unsolved.

III. PROPOSED APPROACH

The authors offer to estimate emotionality based on the users' message analysis. In case when the analysis is performed on a group of users, it involves every message in a group exchange. The analysis employs complex assessments of qualitative and quantitative parameters by using an associative experiment, a method of thematic classification, semantic differentials and ontology.

The authors offer to use "emotional scaling" for the users' message analysis. We will allocate eight emotional scales: S1=«inspiration», S2=«uncertainty», S3=«anger», S4=«discontent», S5=«joy», S6=«fear», S7=«indifference», S8=«disorder». Further, we will adjoin to each scale the corresponding set of semantic differentials received by an associative experiment. The survey has been conducted as follows: users should have provided terms by which they express the feelings specified in scales.

The questionnaire also contained information reflecting social and demographic signs of interviewed users – age, education, gender, marital status. This was due to the hypothesis that associations are variable for demographic groups. Figure 2

depicts the fragment of the questionnaire developed using Google Forms.

Gender:	Marital status:
<input type="radio"/> female	<input type="radio"/> married
<input type="radio"/> male	<input type="radio"/> single
Education:	Age:
<input type="radio"/> secondary education	<input type="radio"/> 16-18
<input type="radio"/> higher education	<input type="radio"/> 19-21
<input type="radio"/> specialized secondary education	<input type="radio"/> 22-25
<input type="radio"/> incomplete higher education	<input type="radio"/> 26-30
<input type="radio"/> bachelor	<input type="radio"/> 31-35
<input type="radio"/> master	<input type="radio"/> 36-40
<input type="radio"/> candidate of sciences	<input type="radio"/> 41-45
<input type="radio"/> doctor of sciences	<input type="radio"/> 46-50
	<input type="radio"/> 50 years and above

Within two minutes, indicate the words or phrases with which you express a feeling of JOY in a social network.

Fig. 2. The fragment of the questionnaire

Information about the experiment and the invitation to participate in the questionnaire has been distributed among users of OSNs such as VK, "Odnoklassniki", "Facebook".

IV. RESULTS OF THE QUESTIONNAIRE

More than 300 people have participated in the questionnaire. Eight groups have been allocated: G1 - men of 16-18 years old with secondary education; G2 - men of 19-21 years old with incomplete higher education; G3 - men of 22-25 years old with the higher education; G4 - women of 16-18 years old with secondary education; G5 - women of 19-25 years old with unfinished higher education; G6 - women of 22-25 years old with the higher education; G7 - women of 19-25 years old with secondary vocational education; G8 - men and women aged from 25 years old with any education.

At the next stage, the authors created the semantic differential database for each scale. It is an interconnected set of tables representing types of groups, interviewed users and term (word) frequencies [14]. Figures 3 and 4 show the fragment of the table for semantic differentials.

1	Радость (joy)	Злость (Anger)
2	благо (welfare)	агрессия (aggression)
3	самодовольствие (complacency)	алкоголь (alcohol)
4	близоречивость (bias)	апатия (apathy)
5	везение (luck)	атака (attack)
6	веселье (fun)	баррикада (Barricade)
7	восторг (awe)	банды (gangs)
8	взаимопонимание (Mutual understand)	бедствие (trouble)
9	внимание (attention)	беспорядочность (recklessness)
10	восторг (excitement)	бесит (hate)
11	вдохновение (inspiration)	бешенство (fury)
12	восхищение (admiration)	бесит (pain)
13	восторг (delight)	бесит (burn)
14	встреча (meeting)	бесит (scream)
15	взбешенность (gai)	брат (enemy)
16	взбешенность (children)	взбешенность (violence)
17	взбешенность (kids)	взбешенность (yell)
18	доволен (pleased)	гнев (wrath)
19	досада (home)	гнет (oppression)
20	достижение (achievement)	гордость (dignity)
21	дружба (friends)	грабёж (robbery)
22	единица (single)	громоздкий (Terrible)
23	жестокость (cruelty)	избавит (fight)

Fig. 3. The fragment of the semantic differentials table

1	Недовольства (Discontent)	Равнодушие (Indifference)
2	антипатия (antipathy)	апатия (apathy)
3	безобразия (ugliness)	безразличие (inaction)
4	бесит (enrages)	бездушность (callousness)
5	бурчать (mutter)	бездарность (carelessness)
6	возмать (resent)	безразличие (numbness)
7	власть (authority)	бесчувственность (insensibility)
8	вздыхать (groan)	все равно (still)
9	ворчать (grumble)	закономерное (arrogance)
10	вспышка (flare up)	капота (lethargy)
11	гаденыш (bastard)	гордость (dignity)
12	глупый (stupid)	депрессия (depression)
13	гнев (anger)	игнор (ignoring)
14	горе (grief)	изломанность (injury)
15	попрекнуть (discredit)	лени (laziness)
16	посада (shame)	наплевательство (rudeness)
17	путь (horde)	наплевательство (silence)
18	самолет (fly)	наплевательство (spit)
19	заставить (force)	не нужен (no need)
20	зло (evil)	незамешательство (noninterference)
21	истерика (hysteria)	неприкаянность (castigation)
22	колочка (thorn)	неудача (unclinging)
23	конфликт (conflict)	нелюбопытство (uninteresting)
24	косяк (jamb)	непоколебимость (steadfastness)

Fig. 4. The fragment of the semantic differentials table

In some cases to reduce the database size authors have considered stemming terms before lookup.

The results have been structured and reduced to a generalized form by the constructed ontology. The relations between concepts in strings of associations were outlined by descriptive logic as a system of nesting and equivalence axioms. Ontology was created in Protege 4.2. "Thing" is a root class of ontology, classes S_1, \dots, S_8 are lower in the hierarchy, and the strings of associations are below these classes. In each item, the relations of equivalence between concepts in the single list were established. Below there is an example of description for the relations in the ontology.

$$Thing \subseteq S_1 \cup S_2 \cup S_3 \cup S_4 \cup S_5 \cup S_6 \cup S_7 \cup S_8,$$

$$S_1 \subseteq outrage \cup nagging \cup \dots \cup gross,$$

$$S_2 \subseteq impress \cup super \cup faraway over \cup \dots \cup idea,$$

$$impress \equiv super \equiv running.$$

Database size is limited by the minimum term frequency or by the maximum quantity of different terms.

The algorithm for estimating the emotionality parameters in the user vulnerability profile is as follows.

Input:

Emotional scales S_1, \dots, S_8

Semantic differentials of emotional scales $(s_1^1, \dots, s_{n_1}^1), \dots, (s_1^8, \dots, s_{n_8}^8)$, where n_1, \dots, n_8 – number of differential terms for each scale.

3. Message: $M_1 = (w_1^1, \dots, w_{m_1}^1), \dots, M_k = (w_1^k, \dots, w_{m_k}^k)$, where m_1, \dots, m_k – number of words/terms in messages. Specific messages were received at the time of analysis.

The authors will use the emotional scaling of the users' messages as an indicator of emotionality in the vulnerability profile. The values of this parameter show how often people use term w_j^i and any term from $(s_1^i, \dots, s_{n_i}^i)$ together in relation to the frequency of their usage in separation, that is, the PMI (Pointwise Mutual Information) index [15].

A. The algorithm

Step 1. Generate a message for analysis: $M = M_1 || M_2 || \dots || M_k$, where || represents concatenation. Punctuation and stop-terms without any semantic or emotional content have been removed. After removal we have new message M^* of size $m^* \leq m_1 + m_2 + \dots + m_k$. Renumber the words again in M^* . New sequence will be w_1, w_2, \dots, w_{m^*} .

Step 2. Set $i = 1$ – the scale number from S, emotional indicator from i-th scale will initialize to $\overline{PMI}(i, M^*) = 0$.

Step 3. For each j-th term in M^* ($j = 1 \dots m^*$) calculate

the PMI index $PMI(w_j^i)$ by the formula:

$$PMI(w_j^i) = \log_2 \frac{f(w_j^i, (s_1^i, \dots, s_{n_i}^i))}{f(w_j^i) f(s_1^i, \dots, s_{n_i}^i)}$$

where $f(w_j^i, (s_1^i, \dots, s_{n_i}^i))$ – number of search results for the joint usage of terms w_j^i from M^* with each term from the i-th scale; $f(w_j^i)$ – number of search results for the term w_j^i ; $f(s_1^i, \dots, s_{n_i}^i)$ – number of search results for terms from the i-th scale. Perform normalization: $\overline{PMI}(w_j^i) = \frac{PMI(w_j^i)}{\log_2 f(w_j^i, (s_1^i, \dots, s_{n_i}^i))}$.

Accumulate: $\overline{PMI}(i, M^*) = \overline{PMI}(i, M^*) + \overline{PMI}(w_j^i)$.

Step 4. If the scale is not the last ($i \leq 8$), then $i = i + 1$ and go to step 3, else stop the algorithm.

B. The experiment

Filtering text from punctuation and stop-terms was carried out using the Tomita-parser software application. The user belongs to the group G3 (male, 23 years old with higher education). The search was performed using the "Yandex" search engine API set up to search in the records and comments of social networks such as "My World", "Maxpark", "Twitter", "Vkontakte", "Hi. Ru", "LiveInternet", "Live Journal", "Odnoklassniki", "Facebook".

The query examples:

- to get the value $f(s_1^i, \dots, s_{n_i}^i)$, for the scale «anger», it's necessary to form the query: «http://blogs.yandex.ru/search.rss?text=aggression | recklessness | frenzy | gross | wrath | fight | fury | ... | rage» (| - disjunction (OR) sign), get the RSS code and take the value of the parameter "count";

- to get the value $f(w_j^i)$, for example, for the term «thoughts» it's necessary to form the query: «http://blogs.yandex.ru/search.rss?text=thoughts», get the RSS code and take the value of the parameter "count";

- to get the value $f(w_j^i, (s_1^i, \dots, s_{n_i}^i))$, for example, for the term «thoughts» with a scale of «anger», it's necessary to form the query: «http://blogs.yandex.ru/search.rss?text= thoughts ^ aggression | recklessness | volcano| wrath| fury| ... | rage» (^ - conjunction (AND) sign), get the RSS code and take the value of the parameter "count".

The table fragment for the emotional indicator calculation on a scale of «Fear» () is shown in Figure 5. The table columns are as follows: 1 – scale differential for S_6 («fear»); 2 – term of the compound message for analysis M^* ; 3 – number $f(w_j^i)$ of search results per term query w_j^i ; 4 – number $f(w_j^i, (s_1^i, \dots, s_{n_i}^i))$ of results for joint usage of terms w_j^i from M^* and each word from the j-th scale; 5 – emotional index $PMI(w_j^i)$; 6 - $\log_2 f(w_j^i, (s_1^i, \dots, s_{n_i}^i))$; 7 – normalized emotional indicator of the term $\overline{PMI}(w_j^i)$.

1	2	3	4	5	6	7
Армия (army)	Уважаемый (distinguished)	292059	5082	-27,35	12,31	2,22
Боязнь (fear)	Граждане (citizens)	240943	9620	-26,15	13,23	1,98
Боязнь (terror)	Товарищи (comrades)	204504	3643	-28,68	12,46	2,14
Дрожь (tremor)	Господа (gents)	334033	10220	-26,53	13,92	1,99
Жасть (horror)	Приглашаем (invite)	238461	1883	-28,65	10,72	2,67
Жуть (uncanny)	Вступить (join)	123464	11759	-24,89	13,52	1,84
Зависть (envy)	Группы (group)	462456	9194	-27,15	13,17	2,06
Космос (fright)	Страны (country)	998000	41258	-26,10	15,33	1,71
Космос (fury)	Полканы (loas)	270843	12779	-25,91	13,64	1,90
Кошмар (nightmare)	Поведает (tell)	15789	398	-26,81	8,64	3,10
Мрак (darkness)	Государство (state)	334765	13511	-26,13	13,72	1,90
Мурашки (horripilation)	СССР (USSR)	247327	10678	-26,03	13,38	1,95
Нападение (attack)	Мнение (opinion)	390687	9496	-27,43	13,24	2,07
Неизвестность (unknown)	Аргументы (arguments)	103785	2485	-26,89	11,28	2,38
Неожиданность (unexpected)	Факты (facts)	477906	10876	-26,86	13,41	2,01
Обман (deception)	События (events)	246644	7214	-26,59	12,82	0,07
Оружие (weapon)	Обращаться (hand)	460727	10990	-26,88	13,42	2,00
Ошибка (mistake)	Создание (create)	360707	18190	-26,37	14,23	1,89
Паника (panic)	Группы (group)	428960	9198	-27,04	13,17	2,05
Переживание (experience)	Предложение (proposal)	237331	4832	-27,09	12,28	2,21

Fig. 5. Fragment of the table of calculation of an indicator of user's emotion on a scale «Fear»

The indicator of user's emotion on the scale of «Fear» $\overline{PMI}(6, M^*) = 108.839$. The experimental estimation of emotional indicators on other scales for this user gave the following results:

- «inspiration» $\overline{PMI}(1, M^*) = 78.44$;
- «uncertainty» $\overline{PMI}(2, M^*) = 120.09$;
- «angry» $\overline{PMI}(3, M^*) = 67.54$;
- «discontent» $\overline{PMI}(4, M^*) = 48.33$;
- «happy» $\overline{PMI}(5, M^*) = 87.76$;
- «indifference» $\overline{PMI}(7, M^*) = 61.04$;
- «disorder» $\overline{PMI}(8, M^*) = 96.31$.

The vector of currently received values $V = (G_r, \overline{PMI}(i, M^*), i = \overline{1.8})$ (r – group id) will be in fact the user vulnerability profile.

We can make preliminary conclusions according to the results of experiments:

- the user emotionality indicator for each emotional scale is in the range from 50 to 300;
- as usual for a particular user there is a certain maximum in one or two emotional scales, which changes slightly over time (when messages are added or deleted);
- there are no emotionless users of social networks, i.e., whose emotional indicators for each scale are approximately equal.

V. CONCLUSION

The algorithm for determining the user vulnerability profile, based on the message emotionality estimation, can be applicable for Russian OSNs. With its usage it is possible to assess the degree of vulnerability for an online social network user to information influences with an emotional component in their messages. In addition, it is possible to analyze incoming messages and filter out dangerous ones.

The algorithm has several important advantages: its results do not depend on the field of application, a preliminary expert assessment of users' messages is not required, and by using search engines it can process significant amounts of information, which leads to an increase in the effectiveness of resistance against dangerous information influences.

Acknowledgment

The study was carried out by the authors within the framework of grant 18-07-01109\18 "Algorithms and protocols for assessing and monitoring availability in large-scale telecommunications networks" of the Russian Foundation for Basic Research.

References

- [1] C. Budak, D. Agrawal, A. El Abbadi, "Limiting the spread of misinformation in social networks [Text]", [WWW '11 Proceedings of the 20th international conference on World Wide Web. Hyderabad, India March 28, April 01, 2011, pp. 665-674].
- [2] I. Semenova, Yu. Monakhov, M. Monakhov, "Research of the Dependencies in Variability of Data Reliability Indicators in Information and Telecommunication Systems", [X International IEEE Scientific and Technical Conference "Dynamics of Systems, Mechanisms and Machines" (Dynamics) Conference Dates: 15–17 November, 2016, Omsk]. DOI: 10.1109/Dynamics.2016.7819082 Electronic ISBN: 978-1-5090-4050-6. Print on Demand(PoD) ISBN: 978-1-5090-4051-3
- [3] R. Axelrod, D. William, "Hamilton The Evolution of Cooperation Science", New Series, vol. 211, No. 4489, (Mar. 27, 1981), pp. 1390-1396.
- [4] D.A. Gubanov, D.A. Novikov, A.G. Chkhartishvili, "Social networks: models of informational influence, control and confrontation", M.: Izdat. fiziko-matematicheskoy literatury [in Russian], 2010, p. 228.
- [5] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, B. Bhattacharjee, "Measurement and analysis of online social networks", [7th ACM conference on Internet measurement, 2007, pp. 29-42].
- [6] M.S. Chwe, "Communication and Coordination in Social Network", [Text] Review of Economic Studies, № 67, 2000, pp. 1-16.
- [7] E. Tarnow, "Conformity and Independence in the Large Group [Electronic resource]", Like Water and Vapor, Access mode: URL: <http://cogprints.org/4274/1/LargeGroupOrderTarnow.pdf>.
- [8] W.A. Mason, F.R. Conrey, E.R. Smith, "Situating social influence processes: Dynamic, multidirectional flows of influence within social networks" [Text], Personality and Social Psychology Review, 2007, pp. 279-300, Access mode: <https://doi.org/10.1177/1088868307301032>
- [9] J. Leskovec, L.A. Adamic, B.A. Huberman, "The Dynamics of Viral Marketing" [Text], HP Labs Palo Alto, CA 94304, 2008.
- [10] M. Granovetter, "Threshold Models of Collective Behavior [Text]", American Journal of Sociology, 1978, vol. 83, no. 6, pp. 1420–1443.
- [11] J. Goldenberg, B. Libai, E. Muller, "Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth" [Text], Marketing Letters, 2001, № 2, pp. 11-34.
- [12] D. Zhang, D. Gatica-Perez, S. Bengio, D. Roy, "Learning Influence among Interacting Markov Chains" [Text], Neural Information Processing Systems (NIPS), 2005, pp. 132-141.
- [13] H.W. Hethcote, "The Mathematics of Infectious Diseases" [Text], 2000, pp. 599-653.
- [14] Y.M. Monakhov, I.I. Semenova, M.A. Medvednikova, N.V. Kostina, "Automated method of assessment of psychosemantic profile of social networks users", European Applied Sciences, ORT Publishing, Stuttgart, Germany, 2013, № 3, pp. 46-48.
- [15] A.A. Chetverikov, "Differential measurement of emotional texts with the PMI-IR algorithm", Materialy XXXVIII Mezhdunarodnoy filologicheskoy konferentsii 2009g. Psikholingvistika, SPb.: Fakultet filologii i iskusstv SPbGU [in Russian], 2009, p. 53.
- [16] E. Tarnow, "Conformity and Independence in the Large Group" [Electronic resource], Like Water and Vapor, Access mode: URL:<http://cogprints.org/4274/1/LargeGroupOrderTarnow.pdf>.