2012 Eighth International Conference on Signal Image Technology and Internet Based Systems

Directed Networks of Online Chats: Content-Based Linking and Social Structure

Vladimir Gligorijević
Department of Theoretical Physics
Jožef Stefan Institute
Ljubljana, Slovenia
http://www-f1.ijs.si/∼gligorijevic

Marcin Skowron
Austrian Research Institute
for Artificial Intelligence
1010 Vienna, Austria
http://www.ofai.at/∼marcin.skowron

Bosiljka Tadić
Department of Theoretical Physics
Jožef Stefan Institute
Ljubljana, Slovenia
http://www-f1.ijs.si/∼tadic

***Abstract*—Online chats are recently shown to result in long-term associations among users, represented by a directed weighted network, similar to dialogs in online social networks. We consider the persistent network which emerges from user-to-user communications found in the empirical dataset from IRC `Ubuntu` channel. The structure of these networks is determined by computing topological centrality measures, link correlations and community detection, and by testing validity of the "social ties" hypothesis. To unravel underlying linking mechanisms, we further study type of messages exchanged among users and users with Web bots, and their emotional content, annotated in the texts of messages. We find that the ranking of the users according to the frequency of their messages obeys Zipf's law with a unique exponent for each message type. Furthermore, the specific hierarchical structure of the network with a strong core as well as its social organization are shown to be closely related with the most frequently used message types and the amount of emotional arousal in them.**

I. INTRODUCTION

In recent years quantitative study of human dynamics on the Web has been intensified owing to the abundance of the empirical data. The interactions among users in blogs and some other online communication systems are *mapped onto suitable type of networks* [10], which are then studied by graph theory methods. Combined with physics theory of nonlinear dynamical systems and machine-learning methods of text analysis (see, for instance, the study of `digg.com` in [9]), a rich data structure with high temporal resolution and texts of messages can be studied. In this way it has been revealed that the emotions, communicated in the text of messages in the data play a role in user collective dynamics on diggs [10], blogs [9], forums [4], online games [14] and chats [6]. On the other side, the online social networks can be considered as the largest existing social structures in the world [16]. However, their dynamical structure, representing how the "friendship" links are actually used, can be much different from common social networks [17].

The Internet-Relayed-Chat (IRC) is another form of communication in which no in advance links occur. In our recent study [7] by the analysis of a large dataset from `Ubuntu` channel, we have shown that the online chats can lead to long-term associations among users. The emergent *persistent network of chats* has a hierarchical network organization, in which both emotions and type of messages play a role. In this work we extend the study of the data to further explore how different types of messages, annotated by text analysis, contribute to building the specific organization of the social graph. We analyse topology of the emergent networks in relation to population of different types of messages and test "social ties" hypothesis on them. Further, the network resilience is studied in respect to cutting the links of a given emotional arousal and to removing the whole "central core" of the network.

II. EMPIRICAL DATA

A. Short description of Ubuntu channel activity

The analyzed data-sets include the recordings of interactions of the publicly accessible IRC channels related to the development of the Ubuntu operating system its maintenance, promotion and user support[1]. Specifically, the sets include the annotated and anonymized communication and action logs for two time ranges: 2004/07/05-2006/12/31 and 2008/01/01-2009/12/31. The logs include information on the interactions conducted by users as well as moderators and channel operators, both human and bots.

Typical communication pattern used in this environment includes a prompt exchange of short text messages between the e-community members. The interaction frequently aid the individual users' search for information on software or services, groups discussions on the ongoing projects, but also for conducting open-domain chats. Any user can join and participate in a conversation.

The data transmitted over the Ubuntu-community IRC channels are considered to be in the public domain and continuously logged by the company archiving bots[2]. The acquired data sets enable for the investigation of interaction patterns between e-community members in its different development phases, the detection of events that trigger extraordinary activities and, with the application of the sentiment and affective states detection systems and resources [15] [13], [12], [2], discovering the relations of such events to the textual expressions of participants' affective states and their online interaction and communication patterns.

[1] http://www.ubuntu.com/
[2] https://wiki.ubuntu.com/IRC/TermsOfService

978-0-7695-4911-8/12 $26.00 © 2012 IEEE
DOI 10.1109/SITIS.2012.117

B. Description of the data collection method

The collecting and pre-processing of the data sets included:

1) Retrieval of all the logs for the set time-frame.
2) Anonymization of the logs by substituting the users IDs with a random number references.
3) Removal of spam and substitution of url links and empty lines with a generic "[url-link]" or "[empty-line]" tags, respectively.
4) Annotation of the utterances with a set of sentiment and affective tools and resources which are described in more details in section II-C.

Due to a change in the logging mechanism of the Ubuntu IRC channels, i.e. no channel modes were logged after 2007/10/16, the data-set was divided in two sub-sets [2004-2006] and [2008-2009]. These sets contain consecutive recordings of dialogs and users actions, and in case of 2004-2006 data-set, also information on the changes of IRC modes by moderators.

C. Details of classifiers used for this data set

The emotional content of the utterance was assessed with the SentiStrength classifier [15]. The applied classifier provides two scores for positive and negative content, on the intensity scale from 1 to 5. In particular, the sentiment score depends on the presence of emotion bearing words which are detected and classified based on lexicons of affective word usage. Each word from the lexicons has a value that correspond to the strength of the emotion attached to it, ranging from -5 to 5. Besides the affective words, the classifier considers also syntactic rules like amplification, reduction, negation, and recognizes repetition of letters and exclamation signs as amplifiers. When such a pattern or syntactic rule is discovered, the classifier applies transformation rules and accordingly modifies the sentence scores. The SentStrength has been designed specifically to analyze online data, and is tailored to process language typical for the Internet based communication, e.g., by correcting spelling mistakes, detecting emoticons and accounting for their emotional content.

Further, the Linguistic Inquiry and Word Count dictionary (LIWC) was used to assign a range categories including the affective, cognitive, linguistic, and social processes related ones. See [12] for information on the complete set of categories provided. Similarly Affective Norm for English Words (ANEW) based classifier was used to assess the valence, arousal and dominance of the utterances [2]. In addition to affective categories and dimensions, the provided annotation includes also information on the dialog act classes [13] and number of characters, words, positive and negative emoticons contained in a single utterance.

D. Size and structure of the data

Size: the files contain annotations for all IRC channels: 2004-2006 375 MB; 2008-2009 462 MB.

Structure: the file extension represents the type of the annotation provided, e.g., .sent - sentiment classification, .da - dialog act classification, .liwc - Linguistic Inquiry and Count categories. Internally, the files are formated as follows:

[timestamp] [action] <user1> <user2> (annotations: if action=[U])

Timestamps: if no timestamp was logged, a time range in which an event occurred is provided instead. It is based on the known time-stamps that were recorded before and after an event, i.e. user's, moderator's or channel operator's action.

List of actions: [J] - joining a channel, [L] - leaving from a channel, [KA] - known as - user changes a nickname, [KO] - kicked off - user1 was kicked off channel by user2, [T] - topic set by user, [U] - utterance. [3] Actions include also a range of modes such as e.g., banning of a user from channel, giving operator rights to user, flood protection[4].

Users: when user2 is listed (with the exception of [U] type of action, described above), he is the one conducting an action, e.g. grants operator rights to user1 or bans user1 from a channel. The user IDs for whom an action "user1 known as user2" were discovered, were connected and used consistently for a complete set of channels. There are separate sets of user IDs for each of the data-sets, i.e., user IDs from 2004-2006 set do not match those from 2008-2009 set.

Bots: the bots were identified using information from the channel operator[5]. In the analyzed data-set, linking of bot-names with different nicknames, e.g. "<bot-name> is now known as <different-nickname>" was observed. The nicknames linked in this way were sometimes used by human operators. To avoid a situation where bots' generated utterances or actions are mixed with those of human moderators, the bots' names were interconnected only when both IDs were referring to a known bot-name: "<bot-name1>" is now known as <bot-name2>". There were two bots identified in 2004-2006 data-set and eleven bots in the 2008-2009 data-set.

Moderators and channel operators: the list is generated based on the changes of the modes[6] that can be only conducted by the IRC channel operators[7]. Since the recording of changes of modes are not present in the logs after 2007, the list of operators is provided only for the 2004-2006 data set.

III. IRC Networks with persistent links

For the purpose of this work we consider the data for `Ubuntu` channel for one year period from 2009. For the network analysis, we use only the subset of the data where user-to-user communication is clearly marked. This reduces the size of the data to $N = 85185$ users and 902892 links with total 1327847 messages. Note that in this dataset we have one Web bot (ID="35177") and that bot-to-user communication are

[3] In case of utterances, if information on user2 is provided, then user1 directly addresses him. This communication link was established only if user1 started an utterance with a nickname of another user who was present on a channel on a given day. The user2 nickname must be further followed with a comma or a colon sign. This is in line with the commonly used style when addressing another user on the IRC channel.

[4] An extensive list of modes is presented at http://deoxy.org/chat/unreal.htm.

[5] https://wiki.ubuntu.com/IRC/Bots

[6] http://en.wikipedia.org/wiki/Internet_Relay_Chat#Modes

[7] https://wiki.ubuntu.com/IRC/IrcTeam/OperatorGuidelines

also included. The data are mapped onto a *directed weighted network* with nodes representing users (and the bot) while the weighted link W_{ij} represents total number of messages sent from node $i \rightarrow j$. Furthermore, we are interested in those user associations that *persist over long time* and thus can be considered as a type of "online social networks" [7]. Such *persistent network* is built of users and their links which survive longer than one day after their first appearance.

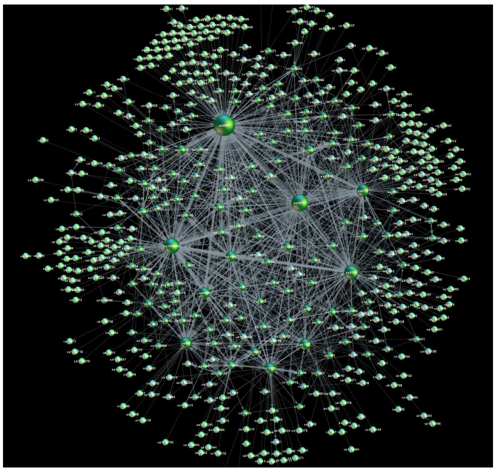

Fig. 1. The persistent network of chats two months after the beginning of the dataset shows characteristic structure with a core, gateway nodes and leaves.

The persistent network has a specific structure, cf. Fig. 1, where the network in its early stage (after two months) is shown. In particular, it builds a strongly connected central core of nodes to which variable number of leaves attach, starting from early stage of the evolution. This leads to hierarchical organization of the network, clearly seen in full size network [7]: the size of the central core increases, at the same time the nodes attached to it acquire cross linking and own leaves, etc. For the entire time period considered (one year) the persistent network appears to have 33838 links among 6168 nodes.

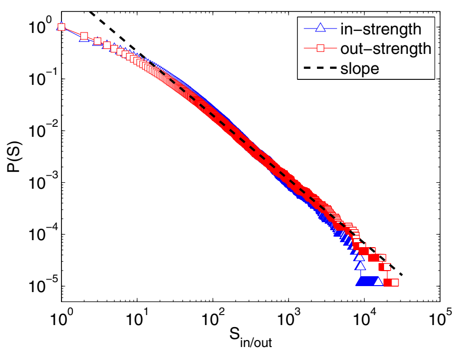

Fig. 2. Cumulative distribution of in-coming and out-going strengths of users in the persistent network of chats, obtained from the data of one year period.

Various measures are determined to characterize the topology of the persistent chats network in our previous Ref. [7]. In particular the scale-free distributions of in-coming and out-going links are found with equal exponents $\tau_{in} \sim \tau_{out} = 2.2 \pm 0.1$ and disassortative mixing with respect to degree (a large-degree node connects to many nodes with small degrees). Here we present additional results of the topology measures for user- and link- heterogeneity in the persistent network.

User strength is topologically well defined quantity $s_i^{out} = \sum_{j=1}^{q_i^{out}} W_{ij}$, and $s_i^{in} = \sum_{j=1}^{q_i^{in}} W_{ij}$, where q_i^{out} and q_i^{in} are out- and in-degree of the node. By construction of the network links, we see that the out- and in-strength measures, respectively, the number of messages sent and received by the user within a given time window. In Fig. 2 we show the distribution of the in- and out-strengths averaged over all users in the persistent network in the one year dataset. The remarkable similarity of the in- and out-strength distributions once again suggests large reciprocity in received and sent messages by a user. The power-law section of the distributions can be fitted by the same exponent $\tau_s = 1.384 \pm 0.008$.

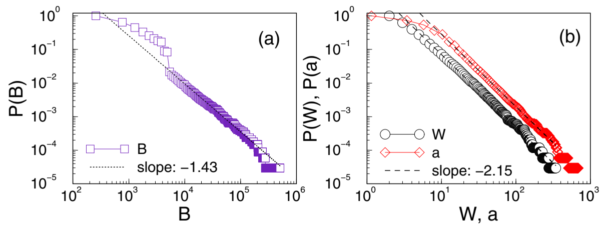

Fig. 3. For the persistent network of chats: (a) Cumulative distributions of the betweenness centrality and of (b) the weight W and the cumulative arousal a carried by the link.

Similar to nodes, the links in the persistent network are heterogeneous, in particular in respect to their *weight* and *centrality* measures. Given that the reciprocity of the links in the network is quite high (cf. Fig. 2 and Ref. [7]), we compute these properties of the links by considering the network as directed. The results are given in Fig. 3a,b. The total weight W_{ij} of the link (i, j) is defined by the number of messages from node i to node j. It represents a *local* property of the network related to the two adjacent nodes. The power-law distribution with the slope $\tau_W = 2.149 \pm 0.009$ of the link weights, shown in Fig. 3b, is a nontrivial feature of this communication system. Similar behavior is found for the distribution of the cumulative arousal a carried by all messages along the link. It departs from the weight distribution at low arousals, mainly because not all messages, which are normally counted in the weight of the link, can be annotated with the emotion classifier algorithm. Thus, their contributions are not included. In contrast, the *betweenness centrality* of a link is a *global* network measure, that quantifies how the link is positioned in the network as a whole. For a given link, betweenness is computed as the number of topologically shortest directed paths among each pair of nodes in the network passing through that link. The distribution of betweenness centrality, averaged over all links in the persistent network, is shown in Fig. 3a. The power-law tail with the slope $\tau_B = 1.43 \pm 0.02$ is another quantitative measure of the persistent network topology.

IV. Structure of the contents-related linking

As explained above in section II, in our data each textual message was analysed to determine the emotional content (arousal and valence) and to annotate the type of message. Here we further analyse (see also [7]) how such additional information about exchanged messages may have contributed to built the particular structure of the user connections.

A. Linking arousal and the network structure

The cumulative arousal that a directed link (i,j) carries is built from all (annotated) messages along that link. In order to test importance of the arousal for the structure of the persistent network, we analyse network resilience with respect to removal of the arousal carrying links. By cutting the links

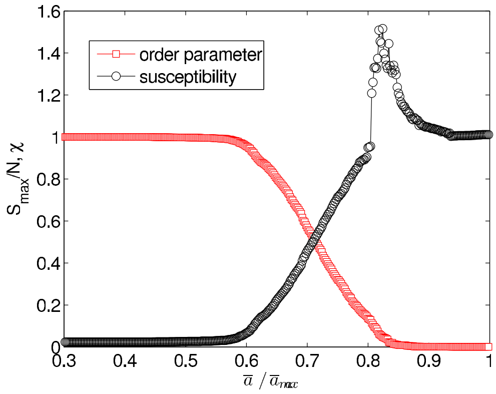

Fig. 4. Relative size S_{max}/N of the giant cluster and fluctuations of cluster sizes χ when the links with the arousal lower than a threshold a are removed, plotted against the scaled threshold a/a_{max}.

which carry the emotional arousal below a threshold $a_{ij} < a$ the giant cluster exists and its size is gradually reduced until considerably high arousal links are cut. The critical arousal where the giant cluster disappears, accompanied by critical fluctuations in the cluster sizes (graph's response function) is about 84% of the maximal link arousal found on the network. This is a kind of bond percolation transition [5], [8], [1] on the graph. Note that different transition point (and the slope) is reached when the links are systematically removed in the increasing order according to the arousal property, as we do here, compared to the decreasing order, as done in [7]. In the latter case the graph starts disappearing from inside the core, while in the former weaker links among nodes at periphery disappear first. Such hysteresis phenomena are familiar to the phase transitions on complex graphs [5].

B. Frequency of message types and the network structure

In addition to the amount of emotional arousal, that is unevenly distributed over the users, the type of messages that a particular user often sends might be characteristic and determine groups of users and their positions in the chat network. Thus the twelve types of messages that are readily determined in our dataset allow further differentiations among users. Having such information in the empirical data, quantitative analysis can be performed. However, there is no standard approach. Here we apply the method that is reminiscent to the analysis of *gene expressions* in bioinformatics [18].

In particular, we first consider frequency of each of the message types in the dataset. We plot ranking distribution of the users who use a given message type ranked according to the message frequency. Interestingly, the use of each message type separately satisfies the Zipf's law, i.e., $f_\kappa(i) \sim r_i(\kappa)^{-\gamma}$ before a cut-off for each type $\kappa = 1, 2, \cdots 12$. The frequency f_κ and the number of users involved n_κ differs among particular message types. Scaled by the average frequency $< f_\kappa >= \sum_i^{n_\kappa} f_\kappa(i)/N_U$, they all fall to a single curve with the slope $\gamma \sim 0.75$, as shown in the main Fig. 5. Moreover, ranking of the message types according to their frequency (popularity among all users) also exhibits a broad distribution, shown in inset to Fig. 5. Originally, the occurrence of Zipf's law in linguistics [3] is related with the frequency of words in a written text and it serves as a measure of syntactic structure of the language. In the present study, the observed power laws are related with how the users are expressed by messages of a given type, i.e., combination of words with specific meaning, such as question, answer, statement, etc. These findings indicate another aspect of users' cooperative behavior on this communication system.

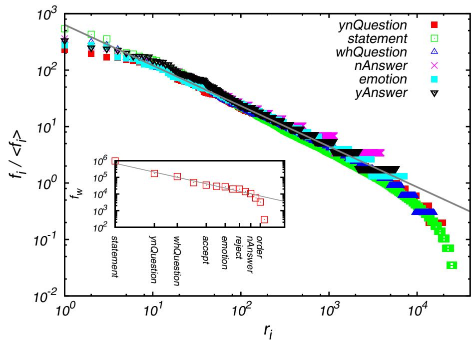

Fig. 5. Scaled ranking distribution of users according to the frequency of used message, for several message types. Inset: Ranking of different message types according to their frequency among all users.

Occurrence of Zipf's law in Fig. 5 suggests high inhomogeneity of the users with respect to the type of messages that they utilize in the communication. In what follows we try to group the users according to their typical type of messages. For this purpose we define a vector $C^i \equiv (c_1^i, c_2^i, \cdots c_{12}^i)$ where each component stands for a particular message type in the data, i.e., $c_1 = ynQuestion$, $c_2 = Statement$, $c_3 = whQuestion$, $c_4 = Reject$, $c_5 = Continuer$, $c_6 = nAnswer$, $c_7 = Emotion$, $c_8 = Accept$, $c_9 = Emphasis$, $c_{10} = Greet$, $c_{11} = yAnswer$, $c_{12} = Order$. Again in the analogy to bioinformatics [18], we assume that a component c_κ is "typical" for an user i if it is expressed by the user more than twice the average for that message type on the channel. Practically, this means that in the ranking distribution in Fig. 5 we draw a line at the level 2 and pick up the user IDs whose ranks are left from the intersection of this line with the ranking distributions. Thus for each user, the vector C^i has units for the components expressed above the threshold line and zeros otherwise.

In this way we obtain how a particular message type is distributed over the users. Notice that a large number of users may have one or very few nonzero components. For instance, the group of users whose vector consists of the first three components (commonly called "question category" message) have a particular attachment to the central core [7]. The most interesting group is the set of users whose vectors consists of all nonzero components. By topological inspection, we find that these users belong to the central core. One of these nodes is the Web bot, whose identity is known from the data. It is interesting that apart from "serving" other users' requests, the users in the central core appear to communicate a lot among themselves and with the bot, cf. density of links in Fig. 6(top).

Other groups with incomplete vector of message types are considered as a whole. In the following, we exclude the central core and all its connections with the rest of the network and consider the remaining network structure. Its giant component consists of 2988 users who are connected with 8031 directed links. View of the network is shown in Fig. 6 (bottom). This network is expected to have different properties, both in terms of connectivity and its "social" structure, which we determine in the remaining part of the section.

In particular, the cumulative distributions of degree and strength of the nodes have smaller exponents ($\tau_q = 0.973 \pm 0.008$ and $\tau_s = 0.866 \pm 0.005$, respectively) and stronger cut-offs compared to the whole network with the core. The removal of the central core and all its links strongly altered the mixing pattern in the remaining network: it shows practically no assortativity, as indicated by the horizontal line in Fig. 7(top), where the average in-degree of node's neighbors is plotted against out-degree of the node. This suggests that the disassortativity found in the whole network [**?**], for comparison shown in the same figure, is a property that originates from the attachment of "poor" nodes to the central core. Furthermore, testing the social "weak tie" hypothesis also shows quantitative changes compared with the whole network when the central core nodes are present. The observed changes are in the direction of (but not equivalent to) recently studied online social structures, e.g., the online social networks [17], [16] and online games [14], where the "weak-tie" hypothesis is found to hold with the universal scaling laws.

As usual, the "weak tie" hypothesis is quantitatively tested by computing how the *overlap* of two adjacent nodes scales with certain property of the link between these nodes, i.e., the width and the betweenness centrality of the link [14], [11]. The overlap O is computed as the normalized number of common neighbors of two nodes which share a given link, and is plotted against the weight of the link W or its betweenness centrality B. In social networks the hypothesis states that two nodes which have strong communications are expected to have large overlap, i.e., the overlap increases with width of the link as $O \sim W^{+\eta_1}$. Similarly, it should decrease with the betweenness of the link as $O \sim B^{-\eta_2}$. In the conventional social networks, the large betweenness links are the bridges between communities. In the online social networks [17] and in the networks emerging from online games [14] it was found that universal scaling laws hold with the exponents close to $\eta_1 = 1/3$ and $\eta_2 = 1/2$, conjectured in [14]. In the online chat networks, however, the absence of typical community structure leads to the scaling laws with different (twice larger) exponents [7]. In the network without central core, that we consider here, we find further changes, which are reported in bottom panel in Fig. 7. Specifically, we find a smaller exponents $\eta_1 = 0.24 \pm 0.10$ and $\eta_2 = 0.69 \pm 0.04$, closer to ones in the online social networks, which can be understood as a consequence of appearance of certain community structure. Indeed, the communities can be visually identified in Fig. 6(bottom). Using the algorithm based on modularity optimization, we can identify about 16 communities of different size, depending on

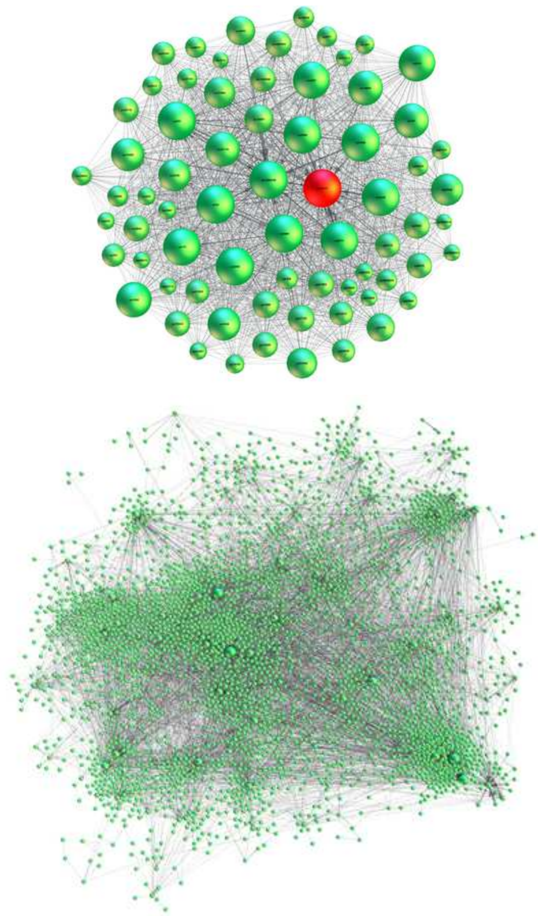

Fig. 6. View of the central core with the bot, marked by different color (top) and the persistent network of chats without central core (bottom).

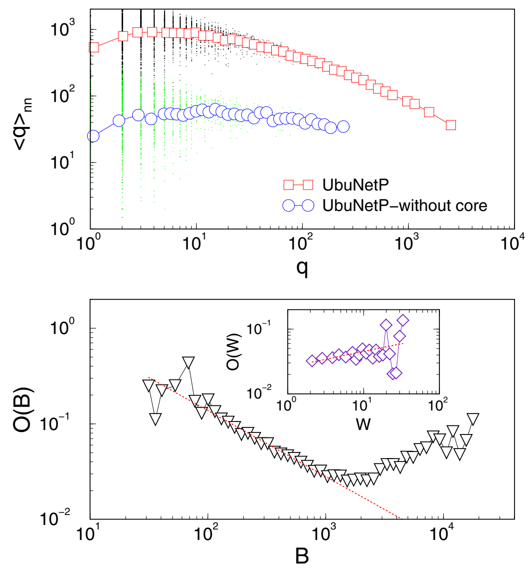

Fig. 7. Top: An assortativity measure computed in the whole persistent network `UbuNetP` and in the network without central core. Bottom: Overlap of pairs of nodes in the network without central core plotted against betweenness centrality B (main figure) and width W of the links between them (inset).

the preset resolution of the algorithm. In the whole network, however, the community structure is absent due to massive interlinking via the central core nodes.

V. Conclusion

Online "social" networks emerge spontaneously in the Internet-Relayed-Chats. Their hierarchical organization, subject to large visibility of the exchanged messages and the roles that certain users assume in the dynamics, is rather different from the structure of `MySpace` dialogs [17]. The social organization in these networks, quantified by the "overlap" measures among users who share a link, depends on the type of messages and the emotional arousal that the messages exchanged along that link are carrying. Although the frequency of different message types is scale invariant (on the channel level), the message types are heterogeneously distributed over users. This affects the social structure of the network. In particular, strong disassortativity and increased overlaps with link weights, found in the network with the central core (consisting of the bot and the users who use all types of messages), changes to virtual absence of these measures when the core is removed. These findings suggest that, despite the content-based linking in the chat channel, the ordinary users away from the central core tend to develop a "social network", with communities and social ties. Further differentiation among users is likely, based on the use of particular words, which can link this quantitative study with ongoing psychology research in user personality profiles. This particular aspect of the data analysis, the role of Web bots and agent-based modeling of online chats are left for future work.

Authors' contributions: BT designed research, analyzed data and wrote the paper; VG contributed software and performed analysis; MS collected data and performed annotation of emotional contents and wrote the related section II.

Acknowledgment

We would like to thank support from the program P1-0044 by the Research agency of the Republic of Slovenia and from the European Community's program FP7-ICT-2008-3 under grant n^o 231323 and COST action MP0801.

References

[1] M. Boguna and M. Angeles Serrano, "Generalized percolation in random directed networks," *Phys. Rev. E*, vol. 72, p. 016106, 2005.

[2] M. Bradley and P. Lang, "Affective norms for English words (anew): Stimuli, instruction manual and affective ratings," The Center for Research in Psychophysiology, Univ. of Florida, Tech. Rep. C-1, 1999.

[3] R. F. i. Cancho and R. V. Sol, "Least effort and the origins of scaling in human language," *Proceedings of the National Academy of Sciences*, vol. 100, no. 3, pp. 788–791, 2003. [Online]. Available: http://www.pnas.org/content/100/3/788.abstract

[4] A. Chmiel, P. Sobkowicz, J. Sienkiewicz, G. Paltoglou, K. Buckley, M. Thelwall, and J. Holyst, "Negative emotions boost user activity at BBC forum," *Physica A*, vol. 390, p. 29362944, 2011.

[5] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "Critical phenomena in complex networks," *Rev. Mod. Phys.*, vol. 80, no. 4, pp. 1275–1335, Oct 2008.

[6] A. Garas, D. Garcia, M. Skowron, and F. Schweitzer, "Emotional persistence in online chatting communities," *Scientific Reports*, vol. 6, 2012.

[7] V. Gligorijević, M. Skowron, and B. Tadić, "Structure and stability of online chat networks built on emotion carrying links," *arxiv:1209.4760*.

[8] A. Kachhvah and N. Gupte, "Transmission of packets on a hierarchical network: Statistics and explosive percolation," *CoRR*, pp. –1–1, 2011.

[9] M. Mitrović, G. Paltoglou, and B. Tadić, "Quantitative analysis of bloggers' collective behavior powered by emotions," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2011, no. 02, pp. P02 005+, February 2011. [Online]. Available: http://dx.doi.org/10.1088/1742-5468/2011/02/P02005

[10] M. Mitrović and B. Tadić, "Emergence and structure of cybercommunities," in *Handbook of Optimization in Complex Networks*, ser. Springer Optimization and Its Applications, M. T. Thai and P. M. Pardalos, Eds. Springer US, 2012, vol. 57, pp. 209–227, [Online] Available: http://dx.doi.org/10.1007/978-1-4614-0754-6$_-$8

[11] J.P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, M. de Menezes, K. Kaski, A.-L. Barabási, and J. Kertész, "Analysis of large-scale weighted networks of one-to-one human communications," *New Journal of Physics*, vol. 9, no. 6, p. 176, 2007.

[12] J. W. Pennebaker, M. E. Francis, and R. K. Booth, *Linguistic Inquiry and Word Count: LIWC 2001*. Erlbaum Publishers, 2001.

[13] M. Skowron and G. Paltoglou, "Affect bartender - affective cues and their application in a conversational agent," in *IEEE Symposium Series on Computational Intelligence 2011, Workshop on Affective Computational Intelligence*. IEEE, 2011.

[14] M. Szell and S. Thurner, "Measuring social dynamics in a massive multiplayer online game," *Social networks*, vol. 3, pp. 313–329, 2010.

[15] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, "Sentiment strength detection in short informal text," *J. Am. Soc. Inf. Sci. Technol.*, vol. 61, pp. 2544–2558, December 2010. [Online]. Available: http://dx.doi.org/10.1002/asi.v61:12

[16] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, "The anatomy of the facebook social graph," *arxiv:1111.4503v1*, 2011.

[17] M. Šuvakov, M. Mitrović, V. Gligorijević, and B. Tadić, "How the online social networks are used: Dialogs-based structure of `MySpace`," *arxiv:1206.6588*.

[18] J. Živkovi, M. Mitrović, and B. Tadić, "Correlation patterns in gene expressions along the cell cycle of yeast," in *Complex Networks*, ser. Studies in Computational Intelligence, S. Fortunato, G. Mangioni, R. Menezes, and V. Nicosia, Eds. Springer Berlin / Heidelberg, 2009, vol. 207, pp. 23–34, 10.1007/978-3-642-01206-8$_-$3. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-01206-8-3