The Behavioral Patterns of Volunteer Computing Communities

Victor I. Tishchenko

FRC "Computer Science and Control RAS"
9 prospekt 60-letya Oktyabrya, Moscow, 117312, Russia
vtichenko@mail.ru

Abstract

The article analyses the model of behavior of the Russian participants of volunteer computing (VC) using platform BOINC. In contrast to the literature data based on sociological surveys our study based on clustering technique shows that the thematic preferences are deceive of the Russian participants of VC motives. And practically there is no any effect on the behavior of a team or individual activity, quantitatively expressed as crédits.

Keywords: volunteer computing, BOINC, virtual communities, complex networks, clustering, bipartite graph

1 Introduction

The complexity of the tasks facing modern science is accompanied by the complexity of the computations necessary to solve these problems. In this case, the idea of such calculations is usually associated with supercomputers or computing clusters. However, not all large-scale tasks require such expensive and specialized equipment.

With the proliferation and rapid development of the Internet, the idea of the founder of the project SETI@home ("Search for Extraterrestrial Intelligence") David Gedye - the use for large-scale distributed computing of the resources of personal computers of Internet users, the integration of which is provided on a special software platform. As such a platform, the BOINC program, developed in 2002 at the University of Berkeley [And], is used.

The methodology of this approach is based on the idea of the network organization of communications and the emergence of the paradigm of the network society as a whole. The principal technical result of solving the problem of parallelizing large-scale calculations and dividing the original problem into many smaller ones was the possibility of organizing ("building") computer systems for voluntary distributed computing (VC) on the principles of grid systems.

Participants in VC projects in accordance with the principles of the national (country) community, thematic or other preferences are combined into teams. The number of 105,582 teams, uniting more than 4 million participants in the VC projects, includes 794 Russian teams. Of 54 active projects using the BOINC platform, four projects operate in Russia: Einstein @ Home, OPTIMA @ home and SAT@home, Gerasim@home. All these projects have the status of "alpha" [Pos14, TP15].

To members of teams and teams as a whole, the organizers of research projects charge conditional points, the so-called "credits", the number of which depends on the provided capacities, the time of participation in projects,

In: E. Ivashko, A. Rumyantsev (eds.): Proceedings of the Third International Conference BOINC:FAST 2017, Petrozavodsk, Russia, August 28 - September 01, 2017, published at http://ceur-ws.org

and other characteristics of the activity of participants. Statistics on the calculation of these indicators, regularly published on the site www.boincstats.com, creates an atmosphere of competition, both between participants and between teams.

This reception organizers of VC projects are considered as one of the incentives for maintaining the activity of volunteers and, accordingly, the motivation for preserving their connection (literally and figuratively) to the project. At the same time, as noted in the literature, in general, the totality of motives for involving Internet users in VC projects and for "connecting" computers to a network of distributed resources, as well as preventing the release of volunteers from projects, still requires its solution [DC10].

The study of the causes of cooperation and communication of Internet users in solving scientific problems goes back to the works analyzing the factors shaping and developing the digital "civil science" [iNAA11]. These studies show that "categories of civil science", in other words, the factors that determine the distribution of participants in voluntary computing, depend on the nature and objectives of the projects. And the whole multitude of volunteers is located in the interval, beginning with the "technical" tasks, in which their participation is reduced only to the provision of computer resources for computations (such as the SETI@home project or the Folding @ home project of computer modeling of protein molecule folding) and ending with more complex problems, for the solution of which both collection and analysis of distributed data are required (Stardust@home projects on classification of interstellar dust particles or Galaxy Zoo construction of visual images/images of galaxies). In total, according to researchers, it is possible to describe three types of projects and, accordingly, three distributions of participants.

The revealed differences in the distribution of volunteers underscore the need for a detailed study of the motives for participating in such projects. And, first of all, in VC projects, entry into which does not require participants to do anything other than downloading the BOINC platform and "connecting" computers to a network of distributed resources. Obviously, the implementation of such projects directly depends on the motives for the participation of volunteers in the project, and, accordingly, on the number of "connected" personal computers and the time of their provision for use in the project. However, the researchers and organizers of VC projects do not have an unambiguous answer to the question: how can a significant scientific project requiring large-scale calculations form an environment that will stimulate the contribution of resources by many volunteers?

According to the researchers of the BOINC community [And14, IK15, HG05, NAA14] the main motivations for participating in the projects of VC participants are:

- a sense of involvement in important scientific research;
- Team spirit, the experience of social interaction, identification with the community, the need for communication with people close to hobbies;
- sports spirit, the atmosphere of competition, the demand for awareness of social status in the form of assessments of social activity (credits).

The basis of the description of these motives and those close to them or related by their characteristics are the results of sociological studies of the VC participants. And naturally they can not be characterized by a high degree of arbitrariness or subjectivity (albeit involuntary) in the evaluation of volunteers' reasons and motives for their participation in VC projects. In this connection, it is important to develop a methodology for verifying the totality of the motives of VC participants based on formalized methods for analyzing their behavior.

2 Statement of the problem. The community BOINC.RU as an integrated network

For a mathematical assessment of the interaction of volunteers, consider the virtual community of Russian VC participants on the BOINC platform in the form of a network. As nodes of the network, we will choose 2 types of objects: community members (user accounts registered on the boinc.ru website) and research projects in which volunteers participate (registered project accounts in the BOINC system). In a graph representing this network, the edge connects one of the vertices belonging to the first type of objects - users, and the other, to the second type - a project in which the user displayed by the first vertex participates (provides its resources for calculations). As a result, we will get a model of the BOINC community network in the form of a bipartite graph with the peaks "participant" and "project". The weight of each rib will be equal to the number of "credits" earned by the participant in the study, with which he is bound by the given edge.

To obtain indicators characterizing the participant of the VC in the BOINC system, the number of points received, and the statistics for each project, sites were used where the indicators are graphically visualized by means of one of the BOINC API applications [BOI].

To conduct a statistical analysis of the behavior of Russian participants in the VC, we also used data obtained with the website www.boinc.ru. The script for obtaining data and creating the corresponding database from this site was written in PHP, MySQL databases were used to store the data. As a result, the following characteristics were obtained: unique participant identifiers, participant names, unique project identifiers, project names, number of credits for participants for the last week, month, year and all the time, participants' ownership of projects, unique team identifiers, team names, membership of participants To the teams.

Thus, a database was created containing data on users who indicated Russia as their "affiliation". In the database, we have accumulated indicators for all projects, including archival ones, in which Russian VC participants took part. This allowed us to calculate the indicators that characterize the patterns of participation of Russian participants in projects and BOINC teams.

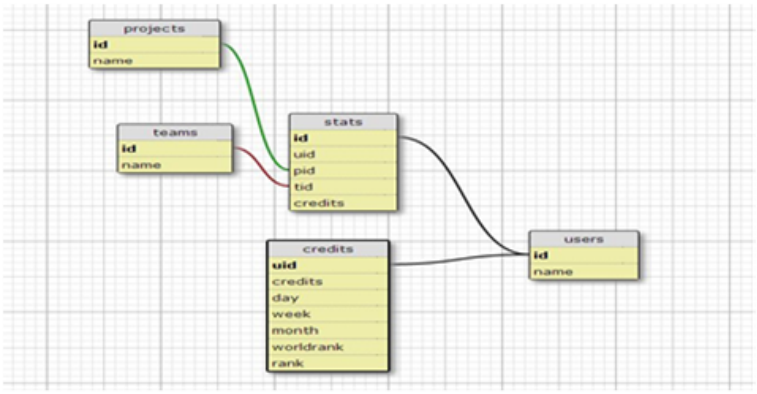

Figure 1: The database structure "BOINC.RU community"

The database "Community BOINC.RU" includes the characteristics of 134 projects and 44985 Russian participants. The data it contains is used to build a network of participants and projects. The network was visualized using the Gephi program. In the constructed bipartite graph, all indicators in the total gave 45119 vertices, 82827 links between them and 794 teams. The average degree of the vertex in the graph is approximately 1.83. The average number of participants in the project is 618. The graph's diameter is 6, the average path length is 2.14.

Figure 2: On the horizontal axis are the projects, on the vertical – the volunteers VC ; the farther the top from the intersection of the axes the higher degree (number of links

3 Analysis of the BOINC.RU community network using clustering methods

As previously noted, as the main reasons for joining the VC projects, researchers based on the analysis of the results of sociological surveys identify three groups of motives - participation in scientific research, team spirit and the need for experiencing competitiveness. To verify these motives, which determine the behavior of the participants in the BOINC.RU community, we clustered the community graph. In our opinion, the results of the selection of clusters can show which of the motives will determine the preference of volunteers in the selection of projects - the scope of scientific interests (subject matter of projects), membership in the team and team participation in the project (team spirit), evaluation of activity (competitiveness), measured by credits.

To assess the significance of the preferences of the participants of the BOINC.RU community, four methods of clustering bipartite graphs were used:

- spectral recursive partitioning method, Spectral Recursive Embedding (SRE) [ZHD+01];
- k-means method [Dhi01];
- The method of partitioning in the main direction, Principal direction divisive partition (PDDP) [Bol98];
- The "bottle neck information" method [ST00].

Algorithms of these methods are developed and applied earlier when clustering collections of documents. Comparison of the results of the methods showed a high degree of applicability of each of them to the network of participants in the BOINC.RU community.

Of the four methods used to cluster bipartite graphs by thematic dependence, only the k-means method did not show any positive results. The methods SRE and PDDP algorithmically distinguished two clusters of participants. The method of the "information bottle neck" actually completely solved the problem of searching for thematic blocks of projects and participants, distributing the entire set of participants on four thematic clusters. Thus, as a result of the formalized analysis, it is shown that the participants' behavior when selecting projects depends on their thematic preferences.

The assumption of the influence of the team spirit on the formation of preferences of community members in selecting projects to which they join was partially confirmed. With available computational resources to bring divisional algorithms to iterations, on which the number of clusters is comparable with the number of real teams, turned out to be an impossible task. For the method using the k-means algorithm, the limiting value turned out to be 14 means, and for the PDDP method with 4 iterations, 16 clusters were allocated. The SRE method proved to be more suitable and allowed to do 5 iterations, breaking the network into 32 clusters. This allowed for the smallest of the resulting clusters to reach the values of the number of participants in the cluster, comparable to the size of the largest of the teams. So, out of the second largest Russian team, Russia Team, consisting of 2,837 people, 89% (2532 participants) joined one of the clusters, consisting of 5 projects and 3,873 participants. All participants of the third largest team "TSC! Russia "were divided into 3 clusters. Since such a strong, almost complete, "entry" is unlikely to be an accident, it should be concluded that belonging to the teams really affects the "project" structure of the community and the preferences of volunteers.

Finally, the idea that the activity of participants, evaluated in the form of loans, will be a strong signal that influences behavior when selecting projects, has not received statistical confirmation. As a result of clustering the graph with all four methods, none of the clusters showed a significant deviation of the average number of credits for all participants. Correlation of the number of loans and the number of projects for all participants was 0.23. In other words, their connection was not strong enough to conclude that there was any influence of competitiveness on the preferences of the BOINC.RU community members when choosing projects.

4 Conclusion

The construction of the mathematical model of the BOINC.RU community as a bipolar graph allowed the clustering by methods previously used primarily for document analysis. As a result, the existing ideas about the motives of the DRV participants' behavior during the selection and accession to the research projects were verified. Unlike the models of behavior described in the literature, participation in scientific research and social interaction, team spirit, are significant. And, in practice, no influence is exerted on the behavior of the participants team or individual activity, the atmosphere of competition, quantified in the form of credits.

The results obtained can have a significant application in solving practical tasks for managing DRV projects and optimizing the work of the participants in the BOINC.RU community.

References

[And] D.P. Anderson. Public computing: Reconnecting people to science. Presented at the Conference on Shared Knowledge and the Web, Madrid, Spain – Nov. 17-19 2003.

[And14] A. Andreev. Metody povyshenya populyarnosti i privlecheniya uchastnikov v proektah dobrovolnyh raspredelennyh vychislenii na platforme (in russian). In *Sbornik tezisov dokladov Nationalnyi Superkompyuternyi Forum (NSCF-2014)*, 2014.

[BOI] Software boinc api applications. http://boinc.berkeley.edu/trac/wiki/GraphicsApi.

[Bol98] Daniel Boley. Principal direction divisive partitioning. *Data Mining and Knowledge Discovery*, 2(4):325–344, Dec 1998.

[DC10] Peter Darch and Annamaria Carusi. Retaining volunteers in volunteer computing projects. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 368(1926):4177–4192, 2010.

[Dhi01] Inderjit S. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD'01*. ACM Press, 2001.

[HG05] Anne Holohan and Anurag Garg. Collaboration online: The example of distributed computing. *Journal of Computer-Mediated Communication*, 10(4):00–00, 2005.

[IK15] V.N. Iakimets and I.I. Kurochkin. Dobrovolnye raspredelennyh vychislenii v rossii (in russian). In *Internet I sovremennoe obschestvo: Sbornik nauchnyh statey XVIII Obyedinenoyi konferentsii (ISM-2015)*, 2015.

[iNAA11] O. *Nov, O. Arazy, and D. Anderson. Technology-mediated citizen science participation: A motivational model. In* Proceedings of the AAAI International Conference on Weblogs and Social Media (ICWSM 2011)*, 2011.*

[NAA14] Oded Nov, Ofer Arazy, and David Anderson. Scientists@home: What drives the quantity and quality of online citizen science participation? PLOS ONE*, 9(4):1–11, 04 2014.*

[Pos14] M.A. Posypkin. Razvitie tehnologii dobrovolnyi raspredelennyh vychislenii v rossii (in russian). In Sbornik tezisov dokladov Nationalnyi Superkompyuternyi Forum (NSCF-2015)*, 2014.*

[ST00] Noam Slonim and Naftali Tishby. Document clustering using word clusters via the information bottleneck method. In Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '00, pages 208–215, New York, NY, USA, 2000. ACM.*

[TP15] V.I. Tishchenko and A.L. Prochko. Rossiiskie uchastniki dobrovolnyh raspredelennyh vychislenii na platforme boinc. statistika uchastiya (in russian). Kompjyternye issledovaniya i modelirovanie*, 7(3):727–734, 2015.*

[ZHD+01] Hongyuan Zha, Xiaofeng He, Chris Ding, Ming Gu, and Horst D. Simon. Bipartite graph partitioning and data clustering. Technical report, may 2001.