ABSTRACTS

The Young Scientist Symposium on Applied Data Analysis (PiterADA 2017)

Galina Pozdniakova
HSE University, St. Petersburg

Abstract title: "Russian political blogs and their subscribers: analysis of the structure of the audience of LiveJournal"

Experts usually associate protests in 2011 all over the world with technological revolution: penetration of the Internet and social networks has reached a critical point, allowing easy mobilization of the masses. (Brancati, 2013). According to Benkler, the online public sphere is a practice of discussing important for society issues which potentially require public recognition and action. Accordingly, internet forums, blogging platforms and social networks become a "meeting place" for users. (Bennett &Segerberg 2012), describe how social networks become coordination centers for mobilizing like-minded people and off-line political movements. As in any other media, each blogger has his own audience, reading his material. (Marlow, 2004) According to Benkler J. (Benkler, 2006) Internet users, are not just passive news readers, but also the creators of content and part of a virtual online community. It is particularly interesting to study political blogs because resulting network clusters can be compared with actual political affiliation of bloggers. The goal of my research is to explore the community of oppositional as bloggers on the LiveJournal platform using SNA (social networks analysis) to describe some possible reasons for inefficiency of Russian opposition.
For this purpose I will calculate all the main metrics of the net. Additionally I tried to answer how the audience of oppositional blogs was divided (or not) by different clusters. Audience activity and popularity of media can be estimated by the two main types of connections. Permalinks looks like references to a specific blog post or an external source in the form of hypertext. Also permalinks are associated with influence of the author of the text on the reader. Blogroll or ties of friendship: adding to the friends create a social bond, later all posts of friends will be shown in a personal blogroll. Regarding blogs, the number of citations and friends, allows to estimate position of the blog in the hierarchy, but the results for both metrics are not always identical.
As a result I received a net of oppositional blogs, colored members according to their political affiliation. Then I compared the theoretical result from clusterization with real affiliations: both distributions differed quite much, also according to common SNA metrics the resulting net had some weaknesses that could have possibly lead to political inefficiency of the non-structured opposition.

References:

1. McKenna, L; Pole, A. What do bloggers do: An average day on an average political blog Public Choice, 2008, 134, 97-108.

2. Wallsten, K. Political Blogs: Transmission belts, soapboxes, mobilizers, or conversation starters J. InfTechnolPolit 2008, 4, 19-40.

3. Marlow, Cameron. "Audience, structure and authority in the weblog community."International Communication Association Conference.Vol. 27. 2004.

4. Benkler, Yochai. The wealth of networks: How social production transforms markets and freedom. Yale University Press, 2006.

5. Pikas, Christina K. "Detecting communities in science blogs." eScience, 2008. eScience'08. IEEE Fourth International Conference on.IEEE, 2008.

6. Zafiropoulos, Kostas, VasilikiVrana, and DimitriosVagianos. "Bloggers' community characteristics and influence within Greek political blogosphere." Future Internet 4.2 (2012): 396-412.

7. Etling, B., et al. "Public discourse in the Russian blogosphere." Mapping RuNet Politics and Mobilization (2010).

8. Vicari S. Exploring the Cuban blogosphere: Discourse networks and informal politics //new media & society. – 2015. – Т. 17. – №. 9. – С. 1492-1512.

9. Park H. W., Thelwall M. Developing network indicators for ideological landscapes from the political blogosphere in South Korea //Journal of Computer-Mediated Communication. – 2008. – Т. 13. – №. 4. – С. 856-879.

10. Haythornthwaite C. Strong, weak, and latent ties and the impact of new media //The information society. – 2002. – Т. 18. – №. 5. – С. 385-401.

11. Chin A., Chignell M. A social hypertext model for finding community in blogs //Proceedings of the seventeenth conference on Hypertext and hypermedia. – ACM, 2006. – С. 11-22.

12. Tremayne M. et al. Issue publics on the web: Applying network theory to the war blogosphere //Journal of Computer-Mediated Communication. – 2006. – Т. 12. – №. 1. – С. 290-310.

Tatiana Merezhko
HSE University, Moscow

Abstract title: "The Shift of Boundaries of Private and Public Spheres: Practices of Mobile Application Usage

Authors: Komissarova Svetlana, Merezhko Tatiana, Mishugina Maria, Sofronova Elena
In the modern world we can easily name a lot of processes which are closely related to the shift of boundaries of public and private spheres. Thus practically everyone owns a smartphone now and consequently can post his/her private data to public sphere at any time. We claim that a spread of new technologies, specifically mobile applications, has an impact on the boundaries of public and private spheres. Our research problem is that there's an empirical discrepancy: Russian smartphone users post their personal data on public social media being not aware of risks to personal life they take.
In our research we based on theoretical interpretation of private and public spheres by U. Habermas, R. Sennet's view on the decline of public sphere and conceptions of «private publicity» and «public privacy» by Z. Papacharissi. In J. Winetraub's works we also found out that socio-cultural characteristics may influence the way different people perceive the shift of boundaries of public and private spheres.
In our research we used mixed design of gathering and analyzing data. Qualitative methodology helped us to reveal motives of publication of personal data, the comprehension of public and private spheres, risk perceptions and opportunities that people see in publication their personal data on the public resources. Quantitative methodology helped us to classify practices of mobile application usage in the context of type of information which is published, and reveal the

essential interrelations using statistics.
During the qualitative stage we carried out 24 semi-structured interviews, each 40 minutes long on the average. The main criterion for the sample was the fact that the respondent use at least one type of mobile applications: social media, beauty and health applications or financial applications. After the results of the interviews had been analyzed, we made a questionnaire considering the results from the qualitative stage. At the quantitative stage we used online survey for gathering data. Proportional sampling was constructed with the data by RLMS-HSE (23th wave), in total 508 respondents were interviewed.
We concluded that the boundaries of private and public spheres lie in the type of information which mobile application users publish on the Internet. They completely oppose posting financial information that is why we can say, that it belongs to the private sphere. On the contrary, various photographs, «superficial» stories and preferences in music and literature, which are also related to personal information by the respondents, belong to the public sphere and can be easily posted. We revealed several types of mobile application users which, according to Z. Papacharissi's approach, are related to the junction of the two spheres, for example, people who are married and have kids, as they are aware of risks they take by publishing information about their family, but use «public private» sphere to create an identity of their family among their friends and relatives. Moreover, we found out that mobile application usage combines convenience and self-control as a smartphone enables a person to have all the functions and constant access to public sphere, and it also confines with a «user agreement» and it requires self-discipline, for example, to keep a food diary.

References:

1. Bodrunova S. Conceptions of Public Sphere and Mediacratic Theory: Searching Touchpoints // Social Communications. URL: URL: http://jourssa.ru/sites/all/files/volumes/2011_1/Bodrunova_%202011_1.pdf

2. Baudrillard J. Ecstasy of Communication // The Anti-Aesthetic. Essays on Postmodern Culture / Ed. H. Foster. Port Townsend: Bay Press, 1983/ URL: http://lib.rin.ru/doc/i/79619p1.html

3. Chen Rui Living a private life in public social networks: an exploration of member self-disclosure // Decision Support Systems. 2012. URL: http://www.sciencedirect.com/science/article/pii/S0167923612003521

4. Cheung A. S.Y. Location privacy: The challenges of mobile service devices // Published by Elsevier. 2013. URL: http://www.sciencedirect.com/science/article/pii/S026736491300201X

5. Freser N. Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy // Social Text. 1990. № 25/26.

6. Habermas J. The structural transformation of the public sphere. Cambridge, Massachusetts: The MIT Press, 1991.

7. Lewis A. Friedland, Thomas Hove, Hernando Rojas. The Networked Public Sphere // Javnost – The Public. 2006. №4.

8. Max van Manen. The Pedagogy of Momus Technologies: Facebook, Privacy, and Online Intimacy // Qualitative Health Research XX(X). 2010.

9. PapacharissiZizi. A Private Sphere: Democracy in a Digital Age / Z. Papacharissi. –

Cambridge, 2010

10. Sennet R. The Fall of Public Man // M: Logos. 2002

11. Weber R. The digital future - a challenge for privacy? // Computer Law & Security Review. 2015. URL: http://www.sciencedirect.com/science/article/pii/S0267364915000047

12. Weintraub J. The Theory and Politics of the Public/Private Distinction // Weintraub Jeff, Kumar Krishan (eds.) Public and Private in Tought and Practice. Perspectives on a Grand Dichotomy. Chicago, London: TheUniversity of ChicagoPress, 1997.

Daria Maglevanaya, Vladimir Yashin, Elene Tabutsadze
HSE University, St. Petersburg

Abstract title:"City as a platform for Web 2.0 resources realization. From bars to Foursquare."

Our project is based on prosumerism which in recent years became an independent form of consumption. Prosumption containing both elements from production and consumption different products (Campbell 2005) also is one of the most important features of web 2.0. Web 2.0 is type of the Internet structure built on users' activity. Hence, culture of prosuming, understood as consumer-generated system, is one of the resources of Web 2.0 (Beer et al. 2010; Shwartz 2012). The aim of our research is to define, how groups from this Internet community transpose on real life structures with characteristic features -- culture preferences. Our question refers to activity in web-application Foursquare and people, who write reviews on places which they had visited. The sample in this project consists of data from Foursquare application. We chose three main bar and restaurant streets in historical center of Saint Petersburg: Rubinstein str., Zhukovskogo str. and Dumskaya str.. Also, we removed from analysis two main avenues of the city as areas with tourist concentration places. Then we selected all places with tags "night life" and "food", because it includes restaurants, bars and other food and drink places, which performs as different capital markers (Bourdieu 1989, Radayev 2016). From this places we took top reviews, which were performed as gate for experts at this application platform.

We assume, that people, who "check-in" in certain type of locations form communities by this cultural preferences that also cause production of public opinions(Cramer 2011; Whyte 1980) . To do this we combine social network analysis and cluster analysis. To explore similarity of agent's reviews lists we use Cophenetic distances. For reviewers community we took criterias of photo includings to their posts, reviews amount, reviewers location (city) (Caquard 2013).
All chosen streets are connected with each other by users' recommendations, but the main differentiation which is contrasting with others is group of those who write reviews, as they concentrate on locations in Rubinshteinastreet.
In future work this project could be used as part of recommendation system for local places of the city as an applied form of next work or it can be used as a better urban lifestyle comparative recognition of global cities in Russia.

References:

1. Beer D., Burrows R. Consumption, Prosumption and Participatory Web Cultures An introduction //Journal of Consumer Culture. – 2010. – Т. 10. – №. 1. – С. 3-12.

2. Bourdieu P. The forms of capital.(1986) //Cultural theory: An anthology. – 2011. – С. 81-93.

3. Campbell, C. (2005) 'The Craft Consumer: Culture, Craft and Consumption in a Postmodern Society', Journal of Consumer Culture 5(1): 23–42.

4. Caquard S. Cartography II Collective cartographies in the social media era //Progress in Human Geography. – 2013. – С. 0309132513514005.

5. Chang J., Sun E. Location 3: How users share and respond to location-based data on social networking sites //Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. – 2011. – С. 74-80.

6. Cramer H., Rost M., Holmquist L. E. Performing a check-in: emerging practices, norms and'conflicts' in location-sharing using foursquare //Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services. – ACM, 2011. – С. 57-66.

7. Cranshaw J. et al. The livehoods project: Utilizing social media to understand the dynamics of a city //International AAAI Conference on Weblogs and Social Media. – 2012. – С. 58.

8. Gehl R. W. The archive and the processor: The internal logic of Web 2.0 //New media & society. – 2011. – Т. 13. – №. 8. – С. 1228-1244.

9. Giurgiu L., Barsan G. The prosumer–core and consequence of the web 2.0 era //Journal of Social Informatics. – 2008. – Т. 9. – С. 53-59.

10. Kesteloot C., Mistiaen P. Brussels: Neighbourhoods as Generators of Integration //Neighbourhoods of Poverty. – Palgrave Macmillan UK, 2006. – С. 198-218.

11. Ritzer G., Jurgenson N. Production, Consumption, Prosumption The nature of capitalism in the age of the digital 'prosumer' //Journal of consumer culture. – 2010. – Т. 10. – №. 1. – С. 13-36.

12. Schwartz R. The networked familiar stranger: An aspect of virtual and local urban anonymity //Seamlessly mobile. – 2012.

13. Toffler A. The Third Wave, chapter The rise of the prosumer. – 1981.

14. Whyte W. H. The social life of small urban spaces. – 1980.

15. Радаев В. В., Котельникова З. В. Изменение структуры потребления алкоголя в контексте государственной алкогольной политики в России //Экономическая политика. – 2016. – Т. 11. – №. 5.

Shafarostov Artem, Soloviev Ilya, Walter Daria, Romov Peter
HSE University, Moscow

Abstract title: "Video-based Monitoring of Audience Attention"

In the current project we develop a framework for video analysis of public presentations and events. Given a video, the aim is to monitor and estimate audience's attention during the event.

We view a video as a sequence of images and process them frame by frame, detecting people's faces, recognizing emotions and classifying people by measured involvement into the presentation.
The work aims to develop and fine-tune the methods of face detection and face classification applied to the our problem. The task we consider is challenging due to a large variability in face size, lighting conditions and occlusions common to real-life videos as well as due to high detection recall required by our problem.
Formalizing and assessing person's interest during a presentation is another difficult task. To solve it, we define an interest index and compute it on each frame using a deep learning approach and face tracking.
Abbreviations: CNN - convolutional neural network
Our problem is divided into three main tasks: face detection, face tracking and face classification. In our project we have made a broad overview of existing solutions in these three tasks. We perform extensive experiments to compare existing methods, libraries and pre-trained neural networks on our particular problem. Face detection is performed using the implementation of Viola-Jones detector from OpenCV and CNN-based methods, for tracking we use Lucas-Kanade algorithm and Kalman filter as well as various geometric heuristics. Emotions, age and gender recognition is currently solved by pre-trained CNNs.
When classifying faces by involvement, we confront the problem of unlabeled data. In future research we are going to cluster the faces using CNN features representations. This approach will let us label the data by groups of involvement and to train our own CNN to classify images.
Our group of researchers use different data obtained from open source (Yandex Machine learning seminars and others) like videos from webinars, university lectures, reports of the company's employees for the analysis of audience interest.
The project is expected to be a pipeline which bundles face detection methods, face tracking and deep learning framework for person involvement classification. Index of person's interest is measured by the means of convolutional neural networks and face tracking methods which allows to estimate face position over time. As an output, for a given video we estimate and visualise the proportion of involved people as well as their average age and gender for each moment of time.
The framework can have multiple real-life applications. For example, it can help assess speaker's skills and qualities, detect the most resonant parts of a presentation, or give a summary of some socio-demographic statistics such as age and gender distribution without conducting surveys.

References:

1.Rasmus Rothe, RaduTimofte, Luc Van Goo, Deep expectation of real and apparent age from a single image without facial landmarks, International Journal of Computer Vision (IJCV), July, 2016

2.Gil Levi and Tal Hassner, Age and Gender Classification using Convolutional Neural Networks, IEEE Workshop on Analysis and Modeling of Faces and Gestures (AMFG), at the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, June 2015

3.Gil Levi and Tal Hassner, Emotion Recognition in the Wild via Convolutional Neural Networks and Mapped Binary Patterns, Proc. ACM International Conference on Multimodal Interaction (ICMI), Seattle, Nov. 2015

4.Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor, Caffe: Convolutional Architecture for Fast Feature Embedding, 2014, http://caffe.berkeleyvision.org

5.OpenCV, Face Detection using Haar Cascades, http://docs.opencv.org/trunk/d7/d8b/tutorial_py_face_detection.html

6.OpenCV, Motion Analysis and Object Tracking, http://docs.opencv.org/2.4/modules/video/doc/motion_analysis_and_object_tracking.html

7.Imagenet classification with deep convolutional neural networks, A Krizhevsky, I Sutskever, GE Hinton. Advances in neural information processing systems, 1097-1105

8.Finding Tiny Faces, Peiyun Hu, Deva Ramanan, arXiv:1612.04402, December, 2016

9.A convolutional neural network cascade for face detection, Haoxiang Li, Zhe Lin, XiaohuiShen, Jonathan Brandt, Gang Hua, Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference

10.FaceNet: A Unified Embedding for Face Recognition and Clustering, Florian Schroff, Dmitry Kalenichenko, James Philbin, arXiv:1503.03832, 2015

11.Microsoft Emotion API, https://www.microsoft.com/cognitive-services/en-us/emotion-api

12.OpenFace: A general-purpose face recognition library with mobile applications, Amos, Brandon and BartoszLudwiczuk and Satyanarayanan, Mahadev, 2016, https://cmusatyalab.github.io/openface/

13.Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks, Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, Yu Qiao, IEEE Signal Processing Letters (SPL), vol. 23, no. 10, pp. 1499-1503, 2016

Nikita Saiapin
Buryat State University

Abstract title: “Modeling of graph structures using Yii2 framework”

Many information systems are using graph structures. We use it for storing information about friends in social networks, users of loyalty programs and so on. However, developers of tree structures face with problem of fast generation of the trees. Many systems are beginning to show a malfunction if the database has a large volume of data. Therefore, the main task of the developers is to develop tree structures, which make all the basic operations faster.

The purpose of work is to implement a system for the simulation of trees using the MySQL database and the PHP-based framework Yii2.
During the work, I developed the database structure for each of the four main methods to store graph structures in databases: «Adjacency List», «Materialized Path», «Nested Sets», and «Closure Table». After that, I developed the program code of these structures using Yii2 framework.

After the implementation of the code, I tested basic tree operations for every developed structure. Therefore, we can make conclusion about every method of storing data in a database:

1. The main advantage of the data storage structure «Adjacency List» is simple implementation of this method. Developers often use this structure for simple operations with a tree, such as

adding, deleting, or moving a node. However, execution of operations that use the analysis of tree is inefficient due to recursion in the program code.

2. The structure «Materialized Path" is well-suited for all major transactions, as well as operations related to the analysis of the tree, but has a major issue associated with the restriction of string which consist the path of the node. Due to this disadvantage, we can build the tree only to several level. In addition, this structure does not allow setting referential integrity, and, as a result, we have to add extra information in every node.

3. The "Nested Sets" structure is excellent when we make request to get full tree or substree. However, the structure is inconvenient to perform basic operations with the trees and does not support referential integrity.

4. The structure of "Closure Table» structure is suitable for all operations, except for getting a subtree for several node, as well as a full tree. These operations require the use of recursive algorithms. Structure has disadvantage of the memory occupied by this structure. This happens because this structure uses two database tables to store information about the nodes of the tree. As a result, I developed four basic tree structures using database and PHP-based Yii2 framework. In addition, I implemented and discussed the advantages and disadvantages of each of the structures.

The system is still finalizing to obtain faster results.

References:

1. Databases. A.Groshev. Archangelsk, 2005.

2. Databases for programmer. S. Tarasov. Moscow, 2015.

3. High Perfomance MySQL. Baron Schwartz, Peter Zaitsev, VadimTkachenko. St.Petersburg, 2010.

4. Introduction to Algorithms, Third Edition. Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein. Moscow, 2013.

5. MySQL Tutorial. Luke Welling, Laura Thomson. Moscow,2005.

6. Official MySQL Tutorial. http://dev.mysql.com/

7.Official PHP Tutorial.http://php.net/

8.PHP Objects, Patterns, and Practice. Matt Zandstra. Moscow, 2011.

9. Simulation hierarchical objects using relational database. D.Ermakov. Yekaterinburg, 2007.

10. Storing Hierarchical Data in a Database. Gijs Van Tulder.
https://www.sitepoint.com/hierarchical-data-database/

11. SQL Antipatterns: Avoiding the Pitfalls of Database Programming. Bill Karwin. Moscow, 2012.

12. Trees and Hierarchies in SQL for Smarties. Joe Celco. 2004.

V. Akimenko, X. Masalygina, V. Nikulichev, I. Osipova, E. Rassochina, L. YalanskiyITMO University

Abstract title: Ensuring data integrity with blockchain technology

Blockchain is a relatively new technology that has shown a lot of possibilities. It emerged in 2009 as a public ledger of all Bitcoin transactions. Blockchain technology is finding applications in wide range of areas: digital assets and stocks, smart contracts, record keeping, ID systems, cloud storage, ride sharing, etc.

The research question was as follows: how to ensure data integrity using blockchain technology?

The following topics are central to our study:

- What is blockchain technology, and how is it typically used?
- What threats does blockchain face? How can such threats be controlled or eliminated?
- What are the key components of data storage security strategy?
- Why does the data storage need to ensure data integrity to be successful?
- How can we use blockchains for integrity assurance?
- What are examples of how blockchain technology is being used?

We investigate the blockchains' activity in terms of how to store, retrieve and share files in decentralized network. We are mostly interested in creating our own version of private chain IaaS "Angelica", that is similar to Storj and BigchainDB, except for the effort of ensuring data integrity in the process. Our model emphasizes the importance of several factors that determine data integrity formulated by Clark and Wilson including well-formed transactions, separation of duty, authentication, audit, Principle of least privilege, objective control and control over privilege transfer.

Our study relies on three whitepapers: Bitcoin, Storj and BigchainDB.

We present and validate a threat model for our solution with mathematical evaluation of variables needed for secure network based on blockchain. We also carry out a brief analysis of possible attackers that are inevitable for our system. This security threat analysis has important significance for revealing the threats that data storages based on blockchain facing. It is used in order to make "Angelica" more secure.

Blockchain technology can secure integrity of files stored in the database. It can be achieved through well-formed transactions, authentication, audit that blockchain provides. The amount of possible threats to data integrity can be decreased.

With one of the three main files' attributes secured blockchain can be used in order to ensure the remaining two properties of data: confidentiality and availability.

There are limitations of using blockchain as it relies on the fact that it is mathematically impossible for a single party to game the system due to lack of needed compute power.

However, with the advent of Quantum Computers, the cryptographic keys may be easy enough to crack through brute force approach within a reasonable time. This will destroy blockchain technology.

References:

1. Clark, D., Wilson, D. A compassion of Commercial and Military Computer Security Policies (1987)

2. Crosby, M., Nachiappan, N., Pattanyak, P., Verma, S., Kalyanaraman, V. BlockChain Technology. Beyond Bitcoin (2015)

3. Lewenberg, Y., Sompolinsky, Y., Zohar, A.: Inclusive block chain protocols. In: Financial Cryptography and Data Security. Springer (2015)

4. McConaghy, T., Marques, R., Muller, A., De Jonghe, D., McConaghy, T., McMullen, G., Henderson, R., Bellemare, S., Granzotto, A. BigchainDB: A Scalable Blockchain Database (2016)

5. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system (2008)

6. Sapirshtein, A., Sompolinsky, Y., Zohar, A. Optimal Selfish Mining Strategies in Bitcoin (2016)

7. Sompolinsky, Y., Zohar, A. Bitcoin's Security Model Revisited (2016)

8. Sompolinsky, Y., Zohar, A. Secure High-Rate Transaction Processing in Bitcoin (2015)

9. Tsirlov, L. Bases of information security of the automated systems.Shortcourse.Phoenix (2008)

10. Wilkinson, S., Boshevski, T., Brando, J., Prestwich, J., Hall, G., Gerbes, P., Hutchins, P., Pollard, C., Buterin V. Storj. A Peer-to-Peer Cloud Storage Network(2016)

11. Wilkinson, S., Lowry J. Metadisk: Blockchain-Based Decentralized File Storage. Application (2014)

12. Zyskind, G., Nathan, O., Pentland, A. Decentralizing privacy: Using blockchain to protect personal data (2015)