Monitoring Obshchestvennogo Mneniya: Ekonomicheskie i Sotsial'nye Peremeny, 2018, vol.145, N3, pages 78-90

Vkontakte' local friendship networks: Identifying the missed residence of users in profile data

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2018 Russian Public Opinion Research Center, VCIOM. All rights reserved. Online social networks (e. g. the most popular Russian website 'VKontakte') are a source of available information about users due to the open data policy. Therefore, researchers have great opportunities to study the topology of interaction networks in the online environment using a social network analysis. However, the personal data that users provide in their public profiles are often incomplete: sections on gender, age or city may be missed inadvertently or skipped intentionally. At the same time, these essential characteristics serve as 'nodes' (i. e. users) and help single out clusters of similar agents and their behavior patterns. The absence of some data can significantly affect network metrics (e. g. size of network, average path length between two participants, distribution of the number of connections between them, etc.) and cause distorted results. In this regard, there is a need to fill gaps in data. The paper presents a case study on the design and applications of a classifier which would determine whether a VKontakte user whose location was not specified in the profile is a resident of a particular city. The classifier was created and tested for the Izhevsk city user network. It is based on the decision tree method which gradually filters the accounts by a series of questions. The paper explains the choice of the main indicators helping the classifier to determine the user's city, describes the algorithm and shows how the network topology changes as the missing data on user's location are added.

http://dx.doi.org/10.14515/monitoring.2018.3.05

Keywords

Big data, Missing data, Network homophily, Network topology, Online communities, Social network analysis, Using R for data analysis, VKontakte

References

[1] Gurin K. E. (2016) Friendship networks structuring of mass media online communities. Discussion. No. 6 (69). P. 64—71. (In Russ.)

[2] Japec L., Kreuter F., Berg M. et al. (2015) AAPOR Report: BIG DATA. February 12, 2015. American Association for Public Opinion Research. Transl. from the English by Rogozin D., Ipatova A., Vyugova E. URL: https://wciom.ru/fileadmin/file/nauka/grusha2015/AAPOR_big_data.pdf (accessed: 1.08.2017). (In Russ.)

[3] Korshunov A., Beloborodov I., Gomzin A. et al. (2013) Detection of demographic attributes of microblog users. In: Proceedings of ISP RAS. Vol. 25. P. 179—194. https://doi.org/10.15514/ISPRAS20132510. (In Russ.)

[4] Chekmyshev O. A., Yashunsky A. D. (2014) Extraction and usage of online social network data. Preprints of Keldysh Institute of Applied Mathematics. No. 62. URL: http://www.keldysh.ru/papers/2014/prep2014_62.pdf (accessed: 1.08.2017). (In Russ.)

[5] Barabasi A. L., Albert R. (1999) Emergence of scaling in random networks. Science. Vol. 286., No. 5439. P. 509—512. https://doi.org/10.1126/science.286.5439.509.

[6] Blondel V. D., Guillaume J.-L., Lambiotte R. et al. (2008) Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. P10008. https://doi.org/10.1088/17425468/2008/10/P10008.

[7] Boyd D. M., Ellison N. B. (2008) Social network sites: Definition, history, and scholarship. Journal of Computer Mediated Communication. Vol. 13. No. 1. P. 210—230. https://doi.org/10.1111/j.10836101.2007.00393.x.

[8] Kosinski M., Graepel T., Stillwell D. (2013) Private traits and attributes are predictable from digital records of human behavior. In: Proceedings of the National Academy of Sciences. P. 5802—5805. https://doi.org/10.1073/pnas.1218772110.

[9] Takhteyev Y., Gruzd A., Wellman B. (2012) Geography of Twitter networks. Social Networks. Vol. 34. No. 1. P. 73—81.

[10] Ugander J., Karrer B., Backstrom L. et al. (2011) The anatomy of the Facebook social graph. Cornell University Library. URL: https://arxiv.org/abs/1111.4503 (accessed: 1.08.2017).