JYX JYVÄSKYLÄN YLIOPISTO UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Semenov, Alexander; Mantzaris, Alexander V.; Nikolaev, Alexander; Veremyev, Alexander; Veijalainen, Jari; Pasiliao, Eduardo L.; Boginski, Vladimir

Title: Exploring social media network landscape of post-Soviet space

Year: 2019

Version: Published version

Copyright: © IEEE, 2019.

Rights: In Copyright

Rights url: http://rightsstatements.org/page/InC/1.0/?language=en

Please cite the original version:

Semenov, A., Mantzaris, A. V., Nikolaev, A., Veremyev, A., Veijalainen, J., Pasiliao, E. L., & Boginski, V. (2019). Exploring social media network landscape of post-Soviet space. IEEE Access, 7, 411-426. https://doi.org/10.1109/ACCESS.2018.2885479

Received August 14, 2018, accepted November 28, 2018, date of publication December 7, 2018, date of current version January 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2885479

Exploring Social Media Network Landscape of Post-Soviet Space

ALEXANDER SEMENOV[1], ALEXANDER V. MANTZARIS[2], ALEXANDER NIKOLAEV[3], ALEXANDER VEREMYEV[4], JARI VEIJALAINEN[1], EDUARDO L. PASILIAO, JR.[5], AND VLADIMIR BOGINSKI[4]

[1]Faculty of Information Technology, University of Jyväskylä, FI-40014 Jyväskylä, Finland
[2]Department of Statistics, University of Central Florida, Orlando, FL 32816, USA
[3]Department of Industrial and Systems Engineering, University at Buffalo, Buffalo, NY 14260, USA
[4] Department of Industrial Engineering and Management Systems, University of Central Florida, Orlando, FL 32816, USA
[5]Air Force Research Laboratory, Eglin AFB, FL 32542, USA

Corresponding author: Alexander Semenov (alexander.v.semenov@jyu.fi)

The work of A. Semenov was supported in part by the Academy of Finland (MineSocMed) under Grant 268078 and in part by the U.S. Air Force Research Laboratory (AFRL) European Office of Aerospace Research and Development under Grant FA9550-17-1-0030. The work of A. Nikolaev was supported in part by the Academy of Finland (MineSocMed) under Grant 268078 and in part by the 2016 Summer Faculty Fellowship through the U.S. Air Force Office of Scientific Research. J. Veijalainen was supported by the Academy of Finland (MineSocMed) under Grant 268078. This work was supported in part by the AFRL Mathematical Modeling and Optimization Institute.

ABSTRACT The "post-Soviet space" consists of countries with a substantial fraction of the world's population; however, unlike many other regions, its social media network landscape is still somewhat under-explored. This paper aims at filling this gap. To this purpose, we use anonymized data on user friendships at VK.com (also known as VKontakte and, informally, as "Russian Facebook"), which is the largest and most popular social media portal in the post-Soviet space with hundreds of millions of user accounts. Using the VK network snapshots from October 2015 to December 2016, we conduct a "multiscale" empirical study of this network by considering connections among individual users, cities, and countries. Our findings indicate that the VK users form a small-world network with basic characteristics consistent with Facebook and other social media networks. In addition, the analysis of modularity-based communities within the user scale network reveals a pattern of geographical separation of the identified communities mostly along the borders between countries. However, the comparison of the two network snapshots suggests that some of these communities may be "blending" within the network, whereas other communities remain "self-contained." Furthermore, the analysis of city scale and country scale networks identifies cities and countries that are most "central" (in the context of certain metrics) in the VK network.

INDEX TERMS Social network services, network theory (graphs), big data applications.

I. INTRODUCTION

Online social media networks have grown enormously over the past decades: some of these networks (i.e., Facebook) nowadays span user populations over the entire globe, whereas others are especially popular among users in certain geographical regions. This paper presents the social network analysis of the largest virtual community in Eastern Europe, based on the anonymized data on user friendships at VK.com (will be referred to simply as "VK" below), formerly known as "VKontakte", which means "in touch" in English. VK is the most popular social media site in Russia, and more generally, in the post-Soviet space. The site was founded in Saint Petersburg, Russia, in 2006 and is often referred to as the "Russian Facebook".

To the best of our knowledge, although VK data has been used as a testbed in some recent studies dealing with social media data analysis, [1] there are no previous studies that analyze the entire VK social network from a global perspective. This paper fills this gap, adding another page to the studies of major online social networks that have so far included Facebook, Twitter, MSN, Weibo, Tumblr, Foursquare, etc. Specifically, we analyze two recent snapshots of the VK social network corresponding to October 2015 and December 2016. In the initially constructed large-scale networks, the nodes represent all individual accounts (also referred to as users) and each pair of nodes is connected by an (undirected) link if the respective users are friends on VK. For the calculations and analysis below, we consider only those users who had

2169-3536 © 2018 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

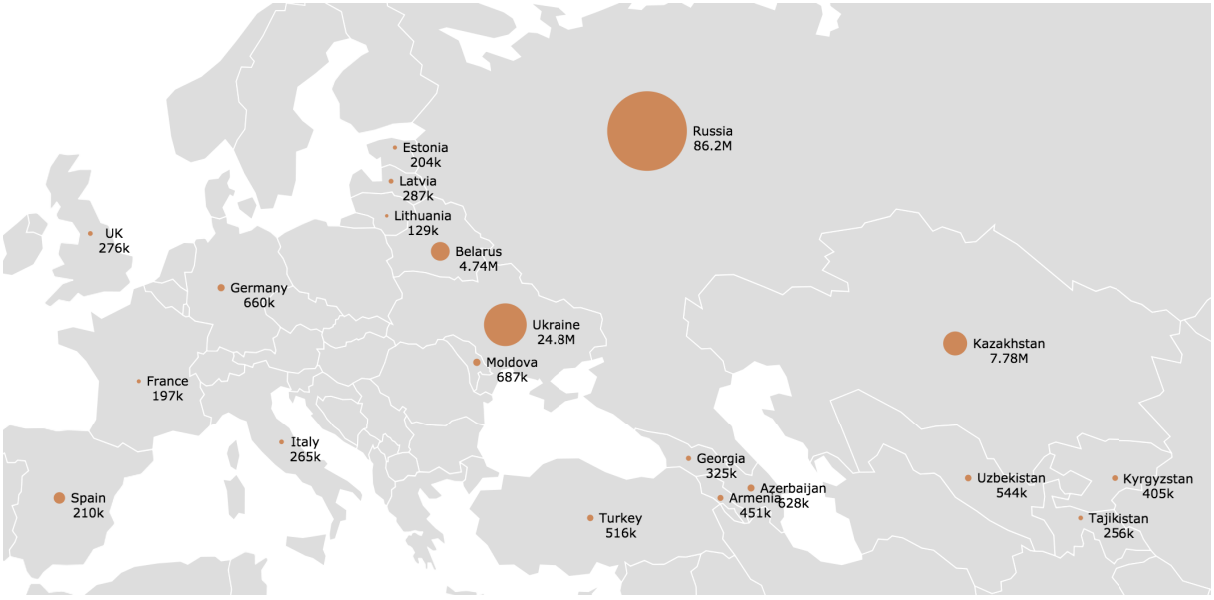

FIGURE 1. A map view of countries with the largest number of VK accounts. The circle area is proportional to the VK population (number of user accounts) as of December 2016 in the corresponding country.

valid (not "banned") VK accounts and who had at least one friend at the time of data collection. Note that we did not impose any restrictions on how recently a user logged in to his/her VK account. The number of VK's registered user accounts was over 300 million for the most recent considered snapshot. Note that some users may have (multiple) duplicate accounts, whereas some accounts may be controlled by link spam bots [2]; however, this work does not aim to explicitly detect or distinguish such accounts. Therefore, all accounts are treated equally and incorporated into the considered networks if they satisfy the aforementioned basic conditions. For simplicity, the terms "user account", "user", and "account" will be used interchangeably throughout this paper.

Moreover, we construct aggregated ("coarsened") networks with nodes representing cities and countries with a substantial number of VK users. Thus, this study deals with a "multiscale" analysis of the VK network, considering graphs that represent individual users, cities, or countries as nodes, as appropriate, and the corresponding connections between these nodes as edges. Mathematically, these networks are treated as undirected graphs, unless otherwise indicated. The broad research questions addressed by this study are:

- What are the "global" connectivity patterns and structural characteristics of the VK network, and how do these properties relate to other previously studied large-scale social media networks?
- Are there any interesting "local" features (i.e., community structures, centrally located nodes) pertinent to certain countries, cities, or regions in the post-Soviet space that can be observed by analyzing the VK network topology at different scales?

We present the respective results in the order of "coarsening" scales of the considered networks. First, we describe basic topological characteristics, as well as modularity-based community structures, of the individual user scale network. Then, we analyze the city scale network, where cities are nodes and links connect pairs of cities that have a sufficiently large number of friendship links between users located in those cities (assuming that each VK user who does share information about his/her location provides the correct information). Finally, we look at the country scale network representing the connections between nodes – major countries – in the post-Soviet space. We then discuss the obtained results in the context of the aforementioned research questions.

II. RESULTS

A. GEOGRAPHY AND DEMOGRAPHICS OF VK

In order to facilitate further presentation and discussion, we briefly summarize basic geographic and demographic characteristics of VK virtual community based on the user profile information. First, we consider the distribution of VK users (with at least one friend) by countries. Fig. 1 illustrates the countries with the largest number of VK users and the corresponding number of users in each country on a map. The total numbers of considered VK users are 141M and 176M for the 2015 and 2016 snapshots, respectively. The user group based in Russia (75.4M and 86.1M VK user accounts in 2015 and 2016, respectively; 144.3M actual population) is by far the largest, followed by the ones in Ukraine (22.4M and 24.7M VK user accounts; 45M population), Kazakhstan (5.87M and 7.78M VK user accounts; 17.8M population), and Belarus (4.26M and 4.74M user accounts; 9.5M population). It is interesting to note that Kazakhstan had the largest relative growth of the number of VK users between the considered data snapshots (32.5%). The USA (not shown on the map) has the largest number of VK users (1.16M and

FIGURE 2. Worldmap of VK user density (regions with the highest concentration of VK users are in red).

1.48M in 2015 and 2016, respectively; 323.4M population) among countries not within the post-Soviet space. The actual population data for the aforementioned countries were taken from respective censuses as of 2016. About 23.9M users did not indicate their country and 31.3M did not indicate their city in the 2015 snapshot. For the 2016 snapshot, 40.6M did not indicate their country, and 50.4M did not indicate their city. These users are not taken into account in the analysis that deals with geographical aspects of the network ''landscape''.

The overall VK user density heatmap is also presented in Fig. 2. As one can observe, Russia and several other post-Soviet countries have a high concentration of VK users, especially in large cities. More details on the distribution of VK users within these countries can be found in Supplementary Materials.

In addition to the geographic aspect, we examined the demographic distribution of VK users by age and sex. The ''under 40'' group represents the majority population of the VK network with the 20-30 being close to half of the number of users, with a slightly higher number of male than female users in this age group. These results are illustrated in more detail in Supplementary Materials. These findings are similar to those in a study performed on the MSN messenger network [3]. However, these results are slightly different from the results on the demographics of Twitter users [4] where a younger (15-24) age range is overrepresented in the network.

B. USER SCALE NETWORK

This section presents the results of VK network analysis at the individual user scale, where nodes represent VK users and two nodes are connected by an edge if the respective users are friends on VK. We calculated several ''global'' topological characteristics of this network, as well as investigated its ''local'' properties, specifically, community structures within this network. The obtained results indicate that the global structural properties of this network are rather similar to the Facebook network, which is not surprising given a similar nature of these networks. Furthermore, the analysis of modularity-based communities that will be discussed later provides interesting insights into the underlying geographical patterns within the VK network.

1) ''GLOBAL'' CHARACTERISTICS OF VK USER SCALE NETWORK

First, we summarize the basic ''global'' structural characteristics of the two snapshots of the VK user scale network, including its size and edge density, degree distribution, clustering coefficient profile, as well as the distribution of the shortest paths, average distance, and diameter.

a: SIZE AND EDGE DENSITY

The first graph representing the snapshot data collected in October 2015 contains 237 million nodes and 5.16 billion edges. Out of these 237 million nodes, roughly 141 million have a degree of at least one, which correspond to VK users who have at least one friend. The graph for December 2016 collection consists of 301 million nodes (out of which approximately 176 million have a degree of at least one) and 6.5 billion edges. As in the case of many social networks, the VK user scale network is very sparse, which is evident from the corresponding edge density (the ratio of the actual to the maximum possible number of edges). The edge density for the 2015 network snapshot is 5.2×10^{-7} and for the 2016 snapshot it is 4.2×10^{-7} after eliminating the isolated nodes. Although the size of the network has substantially increased (the number of nodes has increased

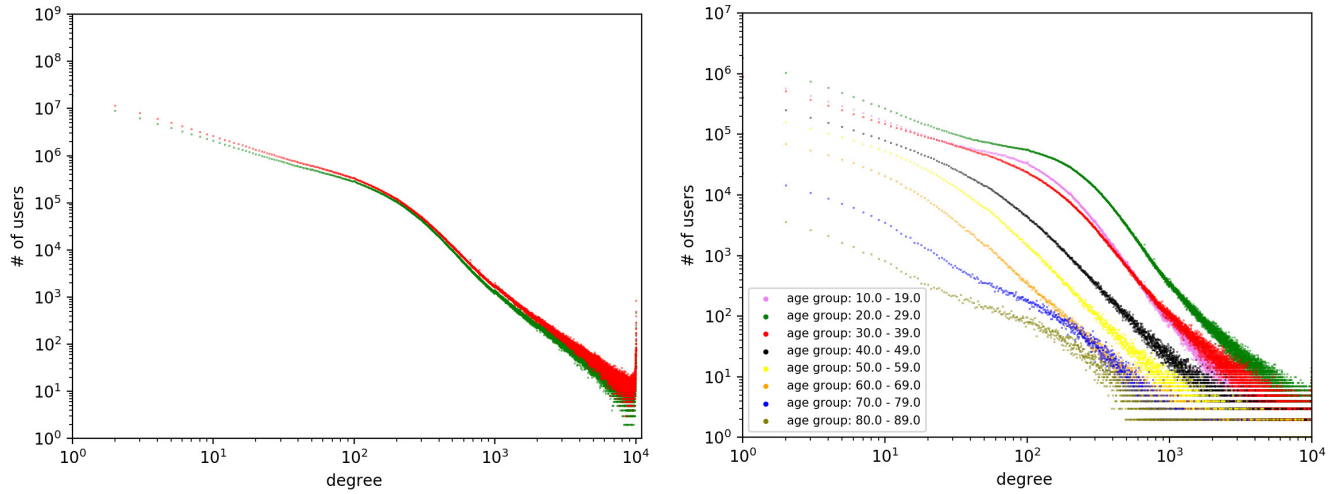

FIGURE 3. Left: VK user scale network degree distributions (logarithmic scale) for 2015 snapshot (green) and 2016 snapshot (red). The inflection point around the degree value of 190 is observed for both snapshots. For the 2015 snapshot: the slope parameters of the power law (along with the respective R^2 values) before and after the inflection point are $\alpha_1 = 0.93$ ($R^2 = 0.99$) and $\alpha_2 = 2.30$ ($R^2 = 0.98$), respectively. For the 2016 snapshot: $\alpha_1 = 0.97$ ($R^2 = 0.99$) and $\alpha_2 = 2.28$ ($R^2 = 0.98$). One can observe that the inflection point is near "Dunbar's number" - the cognitive limit to the number of people with whom an individual is capable of maintaining stable social relationships [5]. Right: Degree distributions of VK user scale network for different age groups (2016 snapshot). The age group of 20-29 years old users is the most active in terms of the number of friends per user, which is consistent with the demographic data of VK users.

by roughly 25%), the network has become sparser as its edge density has decreased by roughly 20%.

b: DEGREE DISTRIBUTION

The degree distributions (logarithmic scale) of the two snapshots of entire VK user scale network is shown in Fig. 3. The degree of each node is simply the number of friends that the respective user has. As one can observe, these distributions are close to a power-law shape; however, there is an inflection point in the degree distribution plot around the degree value of 190, close to the so-called Dunbar's number, [5] which is the hypothesized cognitive limit to the number of people with whom an individual is capable of maintaining stable social relationships. It appears that our data support this hypothesis in the social media domain, since the number of VK users with more than 200 friends decays at a higher rate than the number of users with fewer friends.

In addition, we constructed the degree distribution plots for VK users in different age groups. As it can be seen in Fig. 3, VK users in the age group between 20 and 29 years old are the most active in terms of the number of friends per user; moreover, the aforementioned inflection point occurs at the highest degree value compared to other age groups. This observation is consistent with the fact that this age group is the most represented in the VK user base, as mentioned in the previous section.

c: DEGREE ASSORTATIVITY

As it has been observed in other social networks, the degree of a node's neighbors (friends) is correlated with the node's own degree. That is for smaller-degree nodes, their neighbors' degrees are also small, whereas for larger-degree nodes their neighbors' degrees are also large. This effect is often referred to as *degree assortativity*. To investigate this effect, we calculated average degrees of neighbors for all possible node degrees in the VK user scale network. The results are shown in Fig. 4. The degree correlations are 0.245 and 0.266 for 2015 and 2016 snapshots, respectively, indicating associativity among users and their friends' degrees. For example, the average degree of friends of users with 100 friends is roughly 400, whereas the average degree of friends of users with 300 friends is about 600. The figure also illustrates the distribution of degrees of neighbors of nodes with 50, 100 and 150 neighbors (depicted using violin plots).

d: SHORTEST PATHS AND DIAMETER

Another popular topic of interest in social network analysis is exploring distances (or, shortest path lengths) between the nodes. The length of the longest shortest path between all pairs of nodes is known as the diameter of a network. Due to the very large size of the considered networks, it is almost intractable to calculate the shortest paths between all possible pairs of nodes. Therefore, in order to calculate the shortest paths in the VK network, we have implemented a modified version of the HyperANF algorithm [6] (see "Methods" for more details). In addition to HyperANF, we have sampled 90,000 random (source, destination) pairs of nodes, and computed the shortest paths among those pairs using a breadth-first search algorithm for the 2015 snapshot. The results on the distance distribution are presented in Fig. 5. According to the obtained results, the estimated average shortest path length in the VK network is approximately 4.69 for both of the considered snapshots. This is very close to the average shortest path length of 4.7 for Facebook as of

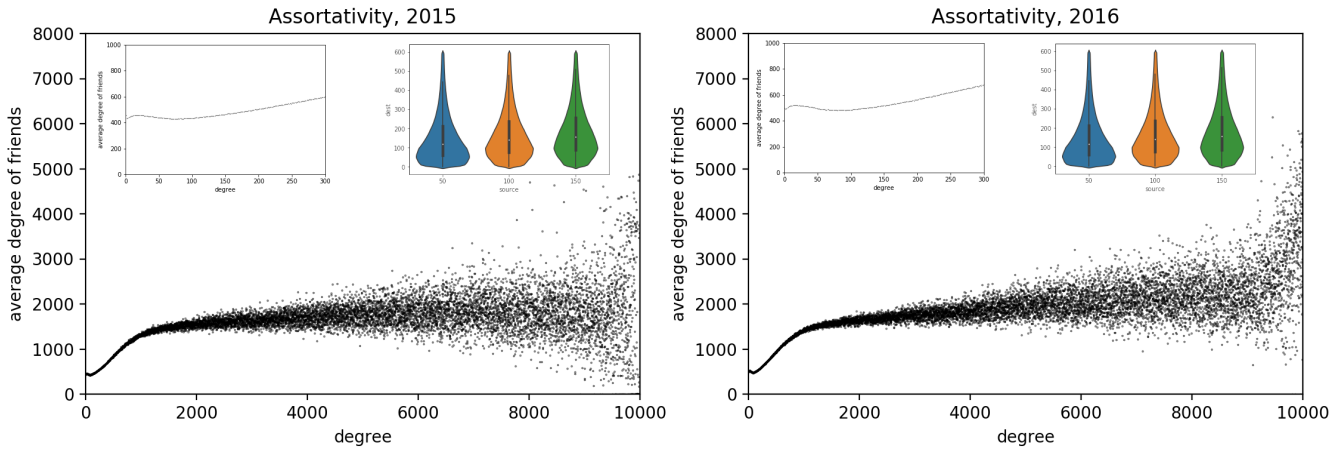

FIGURE 4. The average degree of friends (i.e., degree assortativity) for different node degrees in the VK user scale network: 2015 snapshot (left) and 2016 snapshot (right). The embedded "violin plots" illustrate the distribution of degrees of neighbors for all nodes with 50, 100 and 150 neighbors.

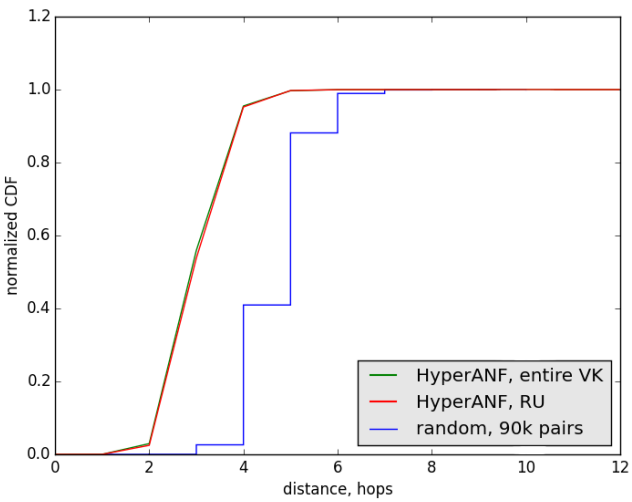

FIGURE 5. The distribution of shortest paths, computed for the entire VK.com network using the modification of HyperANF [6] algorithm, shortest paths within Russia, and results on shortest paths computed for the random samples of 90k source-destination pairs of nodes on the entire VK network for the 2015 snapshot. The results show that most pairs of nodes in VK network are connected by a path of length at most 4 and the average shortest path length is 4.69.

May 2011 [7] or 4.57 reported in a more recent study [8], and generally consistent with the small-world nature of social media networks. The longest shortest path found among the sampled node pairs (which can be treated as the estimate of the graph diameter) was equal to 10.

e: CLUSTERING COEFFICIENT PROFILE

Further, we analyzed clustering coefficients of users in the VK network. As it has been widely reported in the literature, social networks have unexpectedly high clustering coefficients (higher than for a random graph with a similar number of vertices and edges) [9]. Simply put, a given user's friends are more likely to be friends with each other than any other random pair of users. The *local clustering coefficient* of a node i in a network (denoted by C_i) is defined as the ratio of the number of connections among its neighbors to its maximum possible value, that is,

$$C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i}.$$

The average clustering coefficient of a network is the average of all local clustering coefficients calculated for every node i. One can then construct the "clustering coefficient profile", which shows the local and average clustering coefficients depending on the degrees of the nodes. Fig. 6 presents the clustering coefficient profiles of the two network snapshots. As one can observe, the VK network has rather a high clustering coefficients, and the average clustering coefficient of 0.16 is consistent with other online social networks [3], [7].

Overall, the analysis of "global" characteristics of the VK user scale network suggests that it exhibits pronounced small-world properties (low average path length and diameter, high clustering coefficient) that are typical for real-world social networks; moreover, the "bimodal power law" degree distribution with the inflection region around Dunbar's number is another interesting observation.

2) "LOCAL" CHARACTERISTICS OF VK USER SCALE NETWORK: MODULARITY-BASED COMMUNITIES

In order to identify and analyze communities within the VK network of individual users, we employed the modularity-based heuristic community detection algorithm commonly referred to as "Louvain method" [10]. We ran this algorithm on both 2015 and 2016 snapshots of the VK user scale networks and identified over 1,000 communities in each network based on the modularity maximization principle. The resulting modularity values for the 2015 and 2016 snapshots were approximately 0.51 and 0.48, respectively, which indicates that the VK network has an apparent community structure. To our knowledge, these are the largest real-world networks considered in the literature that have been successfully partitioned into modularity-based communities.

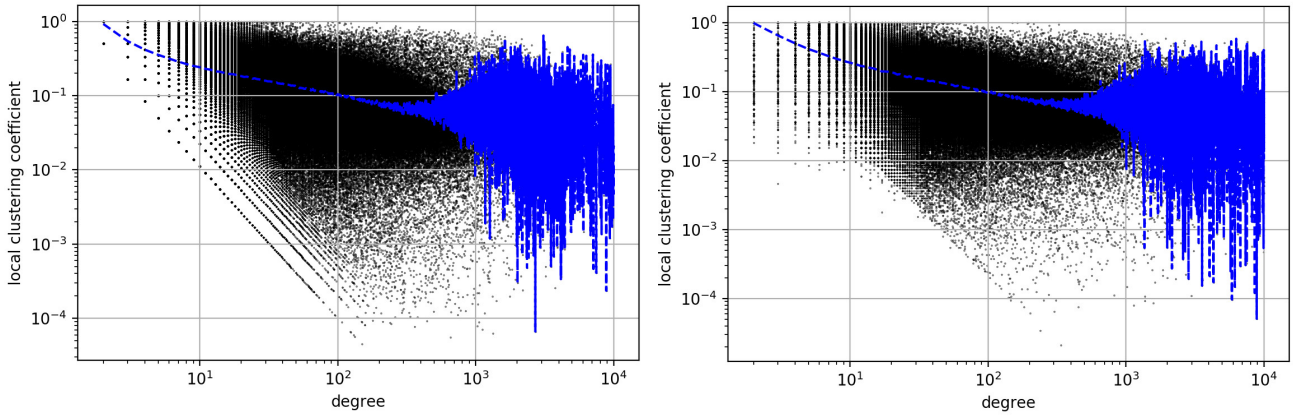

FIGURE 6. Scatter plot of local clustering coefficient to degree (black) and average clustering coefficient to degree (blue). The figures on the left and on the right correspond to 2015 and 2016 network snapshots, respectively. Average clustering coefficient for the entire VK network (both snapshots) is 0.16. Average clustering coefficient for nodes with degree ranging from 2 to 100 is 0.2.

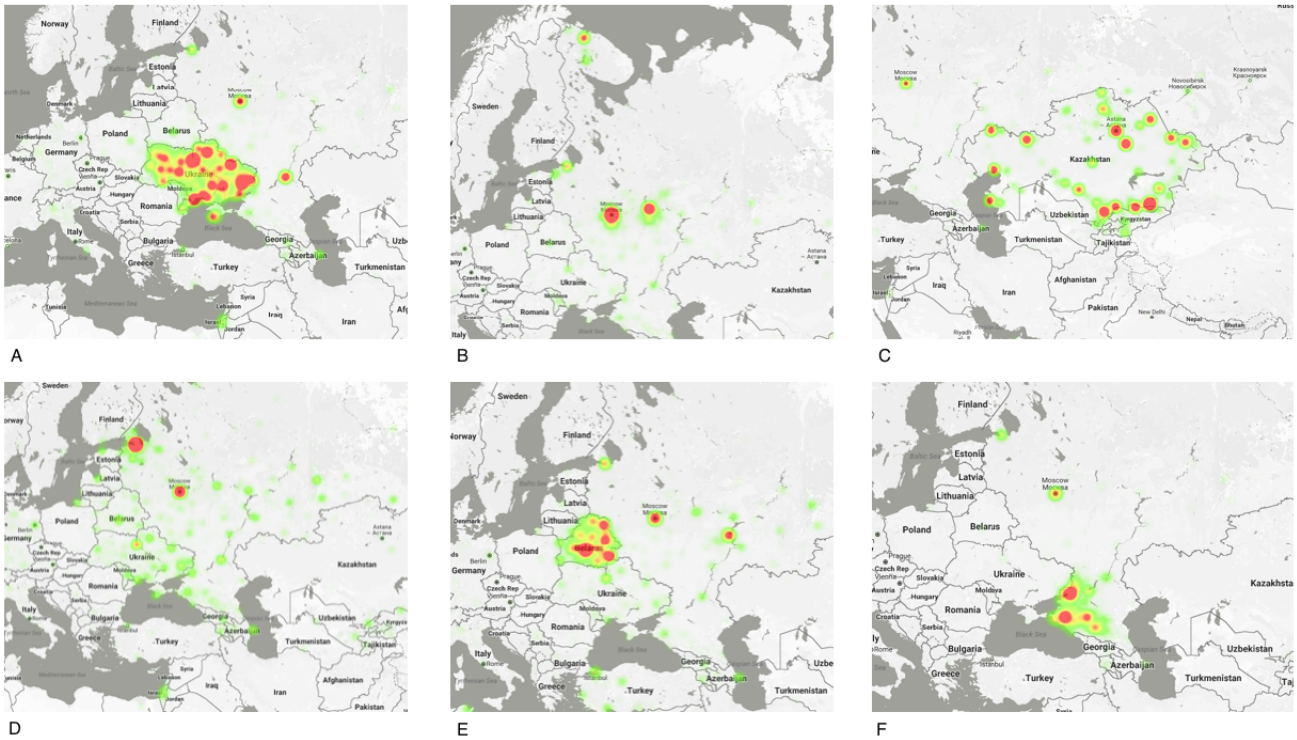

FIGURE 7. Geographic heatmaps of top six largest modularity-based communities in the 2015 VK network of 141M users with at least one friend. (A) The largest community (Ukraine, 20.8M users): conductance = 0.12; (B) 2nd largest community (Moscow & Nizhny Novgorod, 10.8M users), conductance = 0.26; (C) 3rd largest community (Kazakhstan, 8.7M users), conductance = 0.06; (D): 4th largest community (St. Petersburg & Moscow, 7.3M users), conductance = 0.29; (E) 5th largest community (Belarus, 7.2M users), conductance = 0.24; (F) 6th largest community (Southern Russia, 4.8M users), conductance = 0.22.

For the 2015 network snapshot, the algorithm identified 994 distinct communities that contain at least 400 nodes apiece, and approximately 66% of all VK users belong to one of the top 20 largest communities. The largest identified community contains 20.8 million nodes, with a vast majority of the corresponding users located in Ukraine. The size of this largest community is almost twice the size of the second-largest community (approximately 10.8 million nodes; the corresponding users have indicated their location to be mainly Moscow or Nizhny Novgorod region). For the 2016 snapshot, there are 1167 communities having at least 400 nodes.

Interestingly, the trend of observed geospatial concentration of users belonging to the same modularity-based community persists in most of the other identified communities. Fig. 7 illustrates this phenomenon for the top six largest (by the number of nodes) modularity-based

FIGURE 8. Geographic heatmaps of three *new* modularity-based communities in the 2016 VK network. (G) large portion of Russia, 30M users, conductance = 0.22; (H) "industrial cities" in Russia, 5.4M users, conductance = 0.21; (I) Ekaterinburg, Chelyabinsk, and Petropavl (KZ), 5M users, conductance = 0.20.

communities. As one can observe, these communities correspond to Ukraine, Belarus, Kazakhstan, Southern Russia, as well as Moscow, Nizhny Novgorod, and Saint Petersburg, all of which represent prominent regions/cities in the post-Soviet space. It can also be observed that each of the depicted communities contains a significant number of users from Moscow (even in those communities that correspond primarily to other well-defined geographic regions, i.e., Ukraine, Belarus, Kazakhstan, Southern Russia). This is consistent with the fact that Moscow has been growing substantially over the past years, with a lot of people moving to Moscow from various regions/countries, and these people keep social media ties with their places of origin.

We also calculated the *conductance* (also known as *normalized cut*) of each of the identified communities (that is, the ratio of the number of links going outside the community to the number of intra-community links) [11]. Lower values of conductance generally indicate "good" quality of communities in terms of modularity maximization. Fig. 9 presents the summary of the sizes and conductance values of all the identified communities. Among the aforementioned six largest communities, the ones with the lowest conductance values represent Kazakhstan and Ukraine, with conductance equal to 0.06 and 0.12, respectively. This suggests that although the number of VK connections (friendships) in these countries is rather high, these connections are "localized" within the respective countries. The conductance of the community representing Belarus turned out to be higher (0.24), which may be partially explained by tight economic integration of Belarus and Russia. The conductance of the community that primarily contains users in St. Petersburg and Moscow is even higher (0.29), which indicates that this community is connected to many users in other communities.

For the 2016 snapshot, we observe some changes of the properties of the largest communities. It turns out that there are still three large communities corresponding to Ukraine (conductance of 0.14), Kazakhstan (0.11), and Belarus (0.21), although their conductance values changed compared to 2015; moreover, there are three *new* large communities that appear in this network snapshot (see Fig. 8). We examine the aforementioned communities in more detail in the discussion section.

C. CITY SCALE NETWORK

As seen in the previous subsection, geographic locations of VK users do play an important role in both "global" and "local" characteristics of the VK network. Therefore, a natural next step in this analysis is to consider a network that is "aggregated" according to the users' indicated geographic locations, all of which for simplicity are referred to as "cities " regardless of their population sizes. Thus, in this *city scale* network, each city is represented by a node, and edges between nodes represent friendships between individual users in these cities (multiple edges are aggregated into a weighted edge with the weight equal to the number of friendships between the VK users located in the respective cities). In total, there are approximately 250,000 nodes and 32M edges; 16M of those edges had a weight equal to 1. The properties of the city scale network remained virtually unchanged for both 2015 and 2016 network snapshots; therefore, in this subsection we report the results for the 2016 snapshot only.

Further, we have sliced this graph, and removed the edges that have weight less than 1000 (that is, less than 1000 friendships between VK users in respective cities). As a result, we obtained a sliced graph (maximum connected component) with 23,542 nodes and 147,596 edges. The diameter of this (unweighted) graph is equal to 5, the average degree is 5.8, the average shortest path length is 2.27, and the average clustering coefficient is 0.473. Fig. 10 shows the distributions of city sizes (in terms of the number of VK users in a city) and node degrees in the sliced city scale network for the 2016 snapshot. Fig. 11, which is analyzed in detail in the Discussion section, shows the heatmap of the volume of connections between the largest cities (in terms of the number of VK users) in the city scale network. Fig. 12 presents an illustration of the network connecting the largest cities in

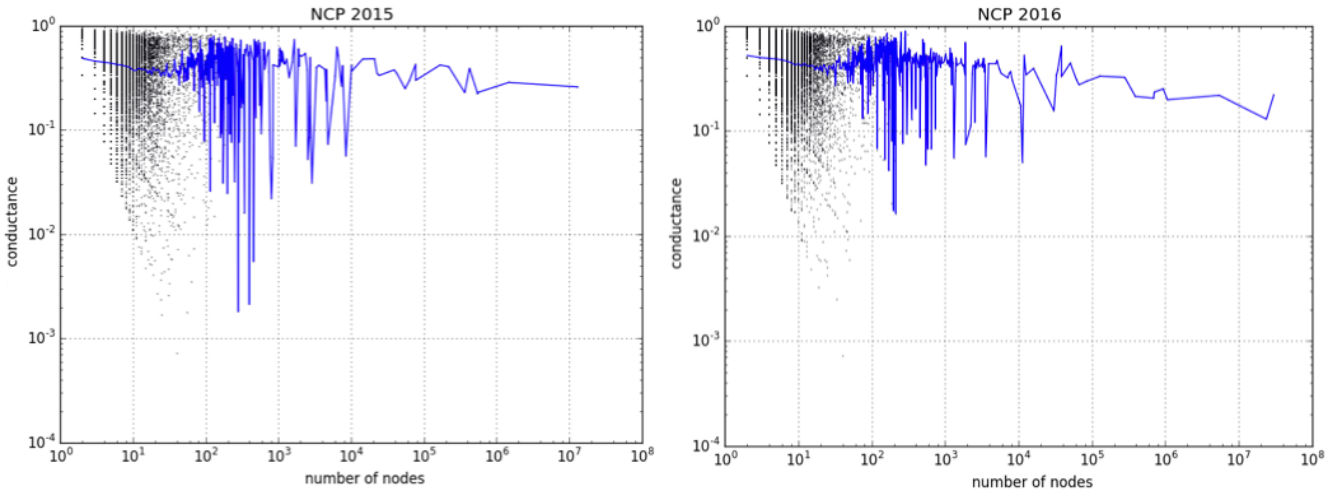

FIGURE 9. Network Community Profiles (NCP) for 2015 (left) and 2016 (right) snapshots of VK network (plots built using a similar approach as in Leskovec *et al.* [11]). The plot shows conductance of communities depending on their size (black: conductance of individual communities, blue: conductance averaged over communities with the same number of nodes). The "best" communities (i.e., the ones with the lowest values of conductance) have around 1,000 nodes, consistent with the results of Leskovec *et al.* [11].

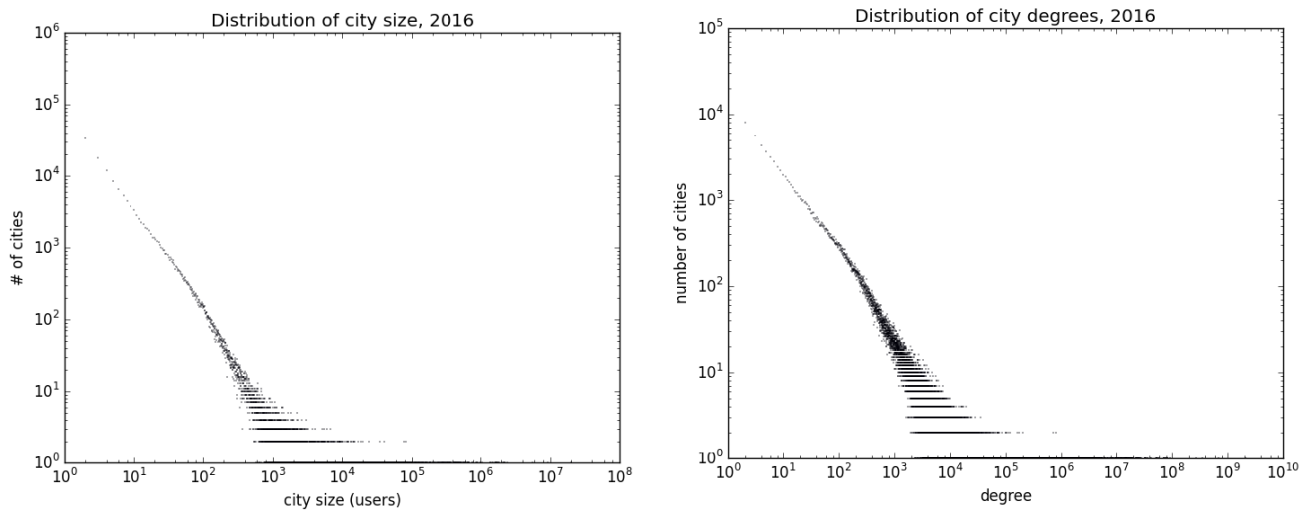

FIGURE 10. Distributions of city sizes (numbers of users) and degrees in the VK city scale network. Degree distribution of the first 70% of nodes ordered by size (cities having degree less than 500) exhibits power law distribution with $\alpha = 0.95$ ($R^2 = 0.98$).

terms of the number of VK users. As one can observe from this figure, most of these cities are in Russia and other post-Soviet countries, with Moscow and St. Petersburg being well-connected to other cities and each other. The two cities not pictured here are New York and Los Angeles. Note that both of these cities have a substantial VK population with ties to Russia and other post-Soviet countries. Another interesting aspect to consider in the context of the city scale VK network is investigating each city's *betweenness centrality* (the score that captures the number of shortest paths between any pair of cities that pass through a specific city). Under the assumption that information diffusion follows shortest paths in social media networks, nodes with high betweenness centrality have a high level of influence on the process of information propagation, rumors, etc. For geographic aggregation (as well as for privacy) reasons, we analyze betweenness centrality of cities rather than individual users. The betweenness centrality of a city is the proportion of shortest paths which go through that city, that is

$$c_B(v) = \frac{2}{n(n-1)} \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

where V is the set of nodes in a network ($|V| = n$), $\sigma(s,t)$ is the number of shortest paths between nodes s and t, $\sigma(s,t|v)$ is the number of such paths passing through the node v. Endpoints were included into shortest path counts. Table 1 presents the top 25 cities in the VK city scale network sorted by their betweenness centrality values. Not surprisingly, the top-ranked cities in this list are the largest cities in the post-Soviet space (Moscow, Kyiv, and Saint Petersburg).

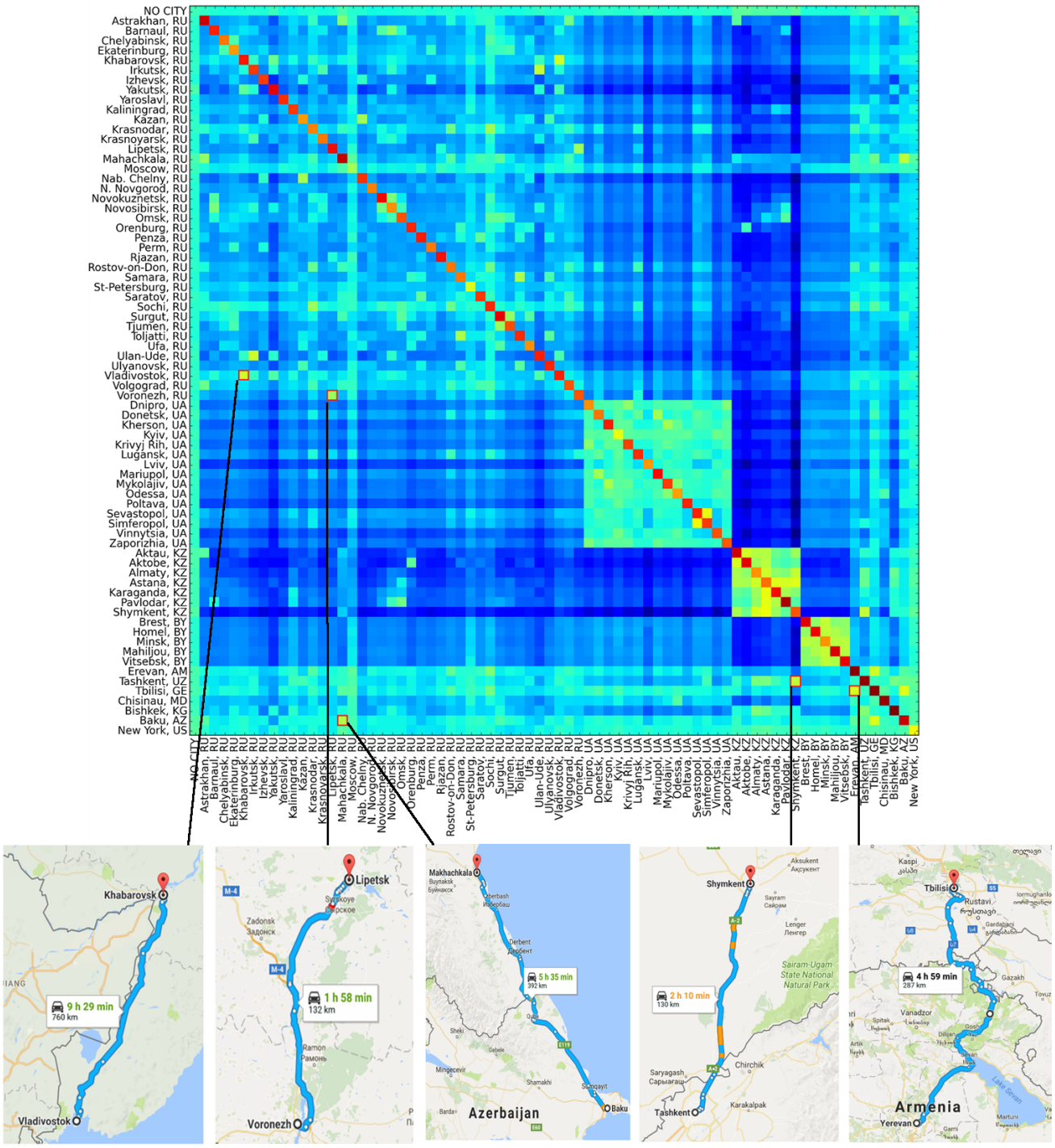

FIGURE 11. **Heatmap: volume of connections between cities in VK city scale network. Cities are grouped by country - left to right, top to bottom: Russia, Ukraine, Kazakhstan, Belarus, other cities with a large number of VK users that are not in these countries. Diagonal elements represent connections within the respective city. Colors in blue spectrum indicate a low volume of connections between the respective cities; colors in the red/yellow spectrum indicate high volume of connections. The heatmap shows a rather "clean" separation by countries, where cities within the same country tend to be more connected. Moreover, cities that are geographically close to each other tend to be more connected. Intercity connections with highlighted squares mark five sample pairs of cities close to each other both "socially" and geographically (Vladivistok-Khabarovsk, Voronezh-Lipetsk, Baku-Makhachkala, Shymkent-Taskhent, Yerevan-Tbilisi). The respective Google Maps distances are shown.**

However, it turns out that the majority of the highest-ranked cities in terms of "betweenness centrality per user" (the ratio of betweenness centrality to the number of users in a given city) are located in Western Ukraine. These cities are highlighted in boldface in Table 1. As one can see from the table, for each of these cities, the value of "betweenness

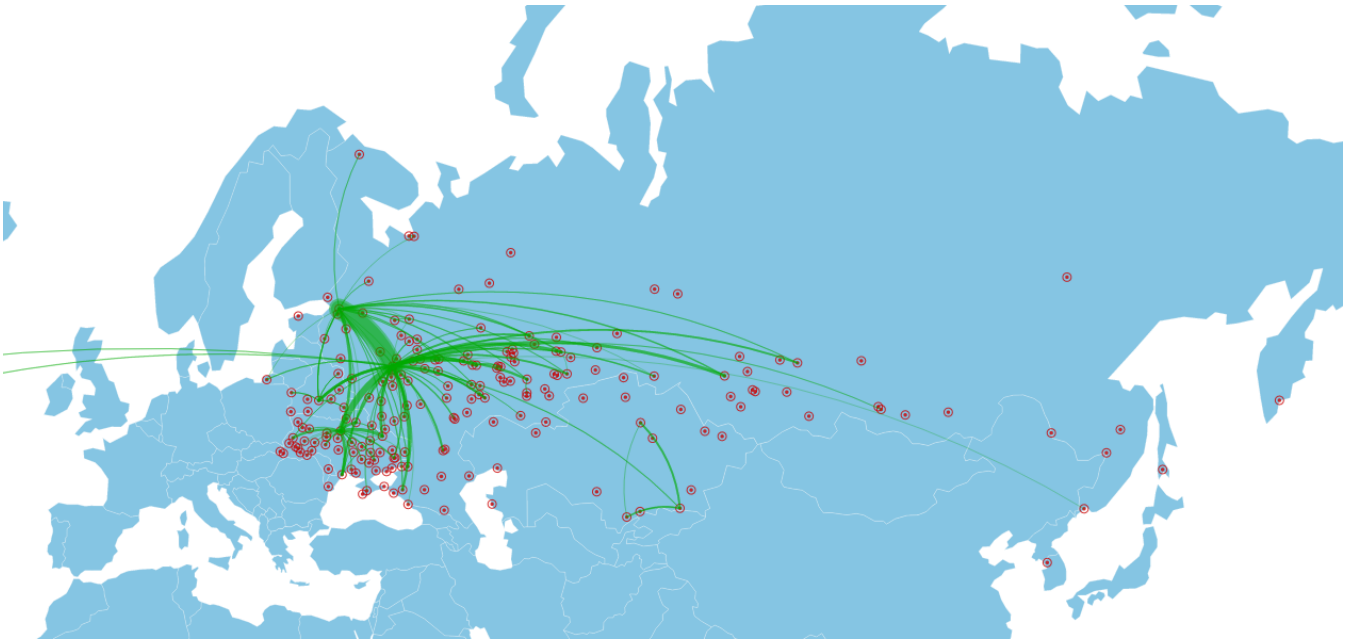

FIGURE 12. Illustration of VK city scale network for the 2015 snapshot. Thicker links represent larger numbers of friendship connections between users in a given pair of cities. A link connecting a pair of cities is depicted if there are at least 1M friendship links between individual users in those cities. Two links that do not have endpoints on the left side of the figure connect Moscow with New York and Los Angeles. The network for the 2016 snapshot is highly similar thus not pictured.

centrality per user'' is larger (in most cases, several times larger) than for any other city on the list. These observations will be addressed in more detail in the Discussion section.

D. COUNTRY SCALE NETWORK

On the most ''coarsened'' scale, we consider the network of countries, where the weights of edges between their nodes are determined by the volume of friendship links between the respective countries, as well as within each country. Due to the small number of nodes in this network (compared to user scale and city scale networks) and the fact that edges exist between all pairs of countries, the analysis of degree distribution, clustering, shortest paths, and communities would be of little practical interest. However, an interesting question to consider for this graph is how users in respective countries are interconnected. To address this question, we construct the matrix of connectivity for post-Soviet countries with the largest number of VK users, which is presented in Table 2. As one may expect, for larger countries, including Russia, Ukraine, Belarus, Kazakhstan, there is a higher volume of intra-country rather than inter-country connections. However, for smaller countries (Armenia, Azerbaijan, Moldova) the connections to Russia are substantially ''stronger'' than those within the respective countries (see a graphical illustration of these results in Supplementary Materials). Thus, on a country scale, Russia clearly plays the most ''central'' role in the VK network.

III. DISCUSSION

In this study, we have described and analyzed the entire social network of VK.com as of October 2015 and December 2016. To our knowledge, this is the first study that deals with the data representing the entire VK network. Due to the high rate of VK.com usage in the post-Soviet space (as it can be seen by a comparison of the number of VK user accounts and the size of the respective actual populations), we believe that this data can provide valuable insights into the social media network landscape of that part of the world. The presented results reveal interesting findings regarding both the overall patterns and the roles of certain regions and countries in ''shaping'' this landscape.

A. ''GLOBAL'' CHARACTERISTICS OF USER AND CITY SCALE NETWORKS

With respect to the ''global'' characteristics of the VK user scale and city scale networks, as one would expect, these networks do exhibit small-world properties with low average path length and diameter, as well as high clustering coefficient. The values of these parameters observed in the VK user scale network are generally consistent with those reported for other planetary-scale social networks, notably Facebook [7], Twitter, [12] and MSN [3]. Table 3 summarizes the basic characteristics of the VK user scale network in comparison with the aforementioned social networks. The degree distribution of the VK user scale network is not a ''pure'' power law, which is consistent with the observations made in the related study of the Facebook [7] and MSN [3] networks, as well as with the recent study that suggested that ''true power law'' networks are rare [13]. More specifically, it turns out that the degree distribution of the VK user scale network can be well approximated by a ''bimodal'' power law model, that is, two power law distributions

TABLE 1. Top 25 cities in the VK city scale network (2015 - table on the left, 2016 - table on the right) sorted by their betweenness centrality. The number of VK users in each city (VK population) and the ratio of betweenness centrality to VK population of each city ("betweenness centrality per user") is also given. Cities with high values of betweenness centrality per user are highlighted in boldface. One can observe that all of the highlighted cities are located in Western Ukraine.

Name	Btw. centr.	VK popul.	Ratio $\times 10^{-7}$
Moscow, RU	0.166	11,410,021	0.145
Kyiv, UA	0.129	3,386,132	0.381
St. Petersburg, RU	0.092	5,806,747	0.158
Lviv, UA	**0.058**	**906,371**	**0.643**
Ufa, RU	0.043	1,122,916	0.381
Kazan, RU	0.036	1,322,779	0.270
Odessa, UA	0.032	1,155,558	0.277
Minsk, BY	0.029	1,346,428	0.215
Vinnytsia, UA	**0.027**	**388,772**	**0.694**
Ekaterinburg, RU	0.025	1,441,260	0.172
Kharkiv, UA	0.025	1,215,465	0.203
Krasnoyarsk, RU	0.024	999,560	0.240
Ternopil, UA	**0.023**	**211,275**	**1.101**
N. Novgorod, RU	0.022	1,059,736	0.205
Chernivtsi, UA	**0.021**	**231,523**	**0.925**
Novosibirsk, RU	0.021	1,389,741	0.152
Perm, RU	0.021	990,427	0.212
Almaty, KZ	0.020	1,323,452	0.151
Ivano-Frankivsk, UA	**0.020**	**204,489**	**0.969**
Dnepr, UA	0.019	1,024,292	0.189
Lutsk, UA	**0.019**	**158,253**	**1.190**
Rivne, UA	**0.018**	**196,816**	**0.936**
Yakutsk, RU	0.018	226,494	0.793
Chelyabinsk, RU	0.018	1,045,146	0.169
Cheboksary, RU	0.017	347,923	0.494

Name	Btw. centr.	VK popul.	Ratio $\times 10^{-7}$
Moscow, RU	0.167	12,994,428	0.130
Kyiv, UA	0.123	3,776,453	0.336
St. Petersburg, RU	0.093	6,533,303	0.142
Lviv, UA	**0.056**	**994,524**	**0.558**
Ufa, RU	0.039	1,255,680	0.309
Kazan, RU	0.033	1,484,251	0.221
Odessa, UA	0.032	1,303,356	0.242
Minsk, BY	0.027	1,494,648	0.180
Vynnitsa, UA	**0.025**	**441,943**	**0.573**
Almaty, KZ	0.025	1,659,469	0.152
Ternopil, UA	**0.025**	**233,538**	**1.049**
Krasnoyarsk, RU	0.024	1,136,225	0.215
Kharkiv, UA	0.024	1,355,813	0.179
Ekaterinburg, RU	0.023	1,644,280	0.138
Chernivtsi, UA	**0.021**	**262,493**	**0.818**
Perm, RU	0.021	1,112,061	0.184
N. Novgorod, RU	0.020	1,195,407	0.169
Ivano-Frankivsk, UA	**0.019**	**228,233**	**0.833**
Novosibirsk, RU	0.019	1,575,561	0.120
Dnepr, UA	0.019	1,137,120	0.163
Rivne, UA	**0.018**	**221,414**	**0.825**
Lutsk, UA	**0.018**	**179,329**	**0.997**
Chelyabinsk, RU	0.017	1,188,528	0,141
Yakutsk, RU	0.016	254,008	0.662
Cheboksary, RU	0.016	384,501	0.426

TABLE 2. Connectivity between post-Soviet countries with the largest number of VK users. Each number in the table is ratio of the number of links between the respective countries (for diagonal elements - the number of links within a given country) to the number of VK users in the country listed in the corresponding row. The presented data corresponds to the 2015 snapshot. The data for the 2016 snapshot is highly similar thus not presented here.

	RU	UA	BY	KZ	AZ	AM	MD
RU	23.52	2.34	0.44	0.23	0.03	0.02	0.03
UA	8.07	27.67	0.46	0.12	0.03	0.02	0.05
BY	8.13	2.47	20.93	0.14	0.02	0.01	0.03
KZ	3.24	0.50	0.10	18.23	0.02	0.01	0.01
AZ	3.18	0.98	0.15	0.19	0.93	0.01	0.01
AM	2.78	0.62	0.10	0.08	0.01	1.03	0.01
MD	3.95	1.46	0.16	0.06	0.01	0.01	3.62

(with different slope parameters on a log-log scale) corresponding to two different ranges of degree values. For both 2015 and 2016 snapshots, we observed that the degree distributions for low-degree nodes (i.e., those with degrees under ∼200, which comprise over 94% of the total number of nodes) follow a power law with the slope parameter just under 1, whereas the degree distributions corresponding to the remaining 6% of higher-degree nodes (over ∼200) also follow a power law but with a substantially higher value of the slope parameter (over 2). An interesting observation is that the inflection point (i.e., the approximate degree value where this change of the decay rate occurs) is just under 200 for both of the considered network snapshots. Moreover, since a very high fraction of VK users (over 94%) have fewer than 200 friends, it suggests that the Dunbar's number phenomenon (the existence of a certain cognitive limit on the number of meaningful social relationships that an individual is capable to maintain, which was estimated at ∼100–200 [5]) may also be observed in the settings of online social media friendships. Our results suggest that the "online version" of Dunbar's number may be slightly higher than the "regular" one, although consistent with the range discussed in [5]. One may intuitively explain a slightly higher number of friendships that people have in the online social media domain by the fact that people sometimes add friends on social media whom they do not know very well in real life; thus, some of such friendships are not as "meaningful" as social relationships maintained in the real life. A related work on Dunbar's number in online social media settings was recently performed using the data from Twitter [14].

Continuing the discussion on the degree distributions, we did observe a significant jump of the fraction of high-degree nodes for degree values close to 10,000. This can be explained by the fact that VK.com imposes the limit of 10,000 friends per user and by the assumption that users who have many friends are trying to achieve the 10,000 marks as this may reflect a certain level of "prestige" in the social media community. Interestingly, both the absolute number and the fraction of VK users with degrees close to 10,000 have substantially increased from 2015 to 2016, which can be seen in Fig. 3. Moreover, as it can be observed from the degree assortativity plots in Fig. 4, the average degree of neighbors (friends) of such high-degree nodes has also substantially increased. Although the number of such nodes is still very small compared to the size of the entire network, it has increased approximately twice from 2015 and 2016; moreover, it appears that high-degree users were adding other

TABLE 3. Summary of basic "global" network characteristics observed in VK and in other major social media networks according to previously published studies.

	VK user scale (2016)*	VK user scale (2015)*	Facebook (2011) [7]	Twitter (2010) [12]**	MSN (2008) [3]
Number of nodes/edges	176M/6.5B	141M/5.16B	721M/68.7B	41.7M/1.47B	180M/1.3B
Average node degree	73.9	73.2	190	**	14
Degree distribution	"bimodal" power-law, inflection point ∼190	"bimodal" power-law inflection point ∼190	non-power-law (not studied further)	out-degree power-law, inflection point $\sim 10^5$	power-law with exponential cutoff
Average distance	4.69	4.69	4.7	4.12	6.6
Diameter (lower bound)* **	10	10	6	18	29
Average clustering coefficient	0.16	0.16	0.14****	Not reported	0.137
Degree correlation	0.266	0.245	0.226	Not reported	Not reported

*Users with at least one friend.
**Note that Twitter network is directed. The average degree needs to be calculated for in-degrees and out-degrees separately.
***The longest distance between a pair of nodes that was reported in the respective study.
****The average clustering coefficient reported for nodes with degree 100.

high-degree users as friends at a higher rate in 2016 than in 2015.

With regard to the degree distribution of the sliced city scale network, we also observed a "bimodal" structure somewhat similar to the one for the user scale network, although only the first ∼70% of the cities (the ones with degrees up to ∼500) follow a power-law degree distribution. The remaining ∼30% of the cities follow a heavy-tail distribution for which a power law is not an appropriate fit. This is due to the fact that there are large cities with very high degrees (i.e., Moscow, St. Petersburg, as well as other large cities in the post-Soviet space). However, interestingly enough, the slope parameter of the power-law part of the city scale network degree distribution is virtually identical to the slope parameter of the user scale network degree distribution before the inflection point. The fact that lower-degree nodes in both the user scale and the city scale networks follow a power-law degree distribution is somewhat expected; however, the fact that the slope parameters of these distributions are almost the same is rather interesting and surprising. In this sense, one may refer to this observation as the *"geographically scale free"* property of the VK network.

B. ANALYSIS OF MODULARITY-BASED COMMUNITIES

Although the aforementioned "global" parameters characterize the behavior of the entire VK user scale network, the description of this network's "landscape" would be incomplete without considering its "local" characteristics, specifically, modularity-based communities within this network. In the discussion of the aforementioned results on these communities, it is important to note that we did not make any *a priori* assumptions about a modular structure of the VK user scale network. Instead, we let the modularity maximization algorithm to "naturally " identify communities within this network and then we mapped these communities onto actual geographic locations of the respective users. Interestingly, most of the largest communities identified by the Louvain algorithm were rather cleanly separated by countries or geographic regions. As one would expect, various regions within Russia, which is by far the most represented country in the VK user base, correspond to some of the largest modularity-based communities. Moreover, there are well-defined communities corresponding to Ukraine, Kazakhstan, and Belarus, which the Louvain algorithm consistently has been able to identify in both network snapshots. It should be noted that each of these communities does contain a substantial number of users located in Moscow, which is understandable due to the fact that people from many regions in the post-Soviet space moved to Moscow.

Since Ukraine, Belarus, and Kazakhstan are among the largest countries in the post-Soviet space and they appear to play a prominent role in shaping a modular structure of the VK user scale network, it is interesting to discuss the properties of the respective communities in more detail and compare their characteristics between 2015 and 2016 snapshots.

As one can observe from the reported results, the conductance of the community corresponding to Kazakhstan is the lowest among all the identified large communities, which implies a relatively "isolated" role of Kazakhstan with respect to the entire social media network of the post-Soviet space. However, it is also worth noting that the conductance of Kazakhstan community has increased almost twice: from 0.06 in 2015 to 0.11 in 2016. This may imply that VK users in Kazakhstan, despite still being relatively "self-contained", are gradually "blending" into the entire VK network and forming more friendship links with users in other countries, most notably in Russia (note that one of the new modularity-based communities that appeared in 2016 spanned users both in Russia and Kazakhstan - see Fig. 8(I)). This observation may also be attributed to an increased rate of migration from Kazakhstan to Russia, although this would need to be confirmed by demographic studies.

Another major modularity-based community corresponds to VK users in Ukraine. In fact, this was the largest community identified in the 2015 VK network snapshot, and it became the second-largest one in the 2016 snapshot after several communities within Russia blended into one large community (see Fig. 8(G)). The conductance of the Ukraine community has stayed almost the same for both snapshots – with a slight increase in 2016, which is not nearly as

significant as in the case of Kazakhstan. As for another major post-Soviet country – Belarus, the respective modularity-based community has a substantially higher conductance value than that of Kazakhstan and Ukraine. For both of the considered network snapshots, the conductance of this community was greater than 0.2, although it slightly decreased from 2015 to 2016. This relatively high conductance value can be explained by historically strong social and economic ties between Belarus and Russia, and our results suggest that these ties to some extent translate to the social media domain.

In addition to the aforementioned large communities that correspond to Kazakhstan, Ukraine, and Belarus, we observed that certain cities and regions within Russia also play an important role in forming large modularity-based communities in the VK network. Moscow and Saint Petersburg are the largest cities in Russia in terms of both actual and VK user population; therefore, it is not surprising that a substantial number of users based in these cities are present in many of the identified large modularity-based communities. In fact, users from Moscow and Saint Petersburg form their own community in the 2015 network snapshot, in addition to being present in most of the other large communities. Interestingly, we observed that several communities that Louvain algorithm identified as distinct (although with rather high conductance values) in the 2015 snapshot have ''blended'' into one extremely large community in the 2016 snapshot (Fig. 8(G)). This observation suggests that Russian VK users overall are getting more connected with each other and the effects of geographical separation on forming friendship links between users within Russia may be gradually diminishing. It should also be noted that a relatively large community identified in the 2016 network snapshot (Fig. 8(H)) corresponds to cities that are not in close geographic proximity, but are somewhat similar due to another criterion: the fact that there are large plants and factories from various types of industry. Our results suggest that VK users in these ''industrial'' cities have formed significant social media friendship ties, which may be related to business interactions between industries in these cities. Therefore, although the geographic proximity factor still appears to be predominant in shaping the modular structure of the VK network (as the analysis of VK city scale network also suggests), there are other factors that may also be important in forming such communities. In particular, it should be noted that in addition to communities formed based on geographic principles, VK users (as well as users of other social networks) also form communities based on their common interests. VK.com allows creation of such interest groups, and it is typical that users that join a particular group become friends with each other. However, communities formed by such users would typically be much smaller than the ones corresponding to geographic regions discussed above. Although in this study we did not track the users' memberships in interest groups in VK, it would be an interesting topic for future studies to analyze the behavior of smaller-size communities and investigate whether the ''best'' communities (in the sense of low conductance, as illustrated in Fig. 9) are formed primarily as a result of geographic proximities, memberships in common interest groups, or other factors.

C. ANALYSIS OF CITY SCALE AND COUNTRY SCALE NETWORK CONNECTIVITY PATTERNS

Shifting the discussion focus to the ''coarsened'' scale networks (city scale and country scale), we note that although some information that is present in the user scale network is inevitably lost after aggregating individual user nodes and edges, the analysis of these aggregated networks helps reveal new information on the collective behavior of VK users in certain cities, regions, and countries. As discussed above, the ''global'' characteristics of the city scale network are similar to those of the user scale network, which motivated us to use the ''geographically scale free'' metaphor to describe these properties of the VK network. However, a more practical aspect of this analysis is identifying cities that are ''central'' with respect to the landscape of the considered networks. In this study, we used betweenness centralities to measure the roles of specific cities in the city scale networks. Although betweenness centrality is obviously not the only way to measure ''central'' roles of nodes in a network, we believe that this is a practically relevant approach in the considered settings, since it measures the fractions of shortest paths ''controlled'' by users in the respective cities. Under the assumption that information diffusion on social media networks is likely to follow shortest paths, users in these ''central'' cities will be exposed to more information reposting cascades (and potentially generate more information cascades that will be seen by many other users) than users in cities that are ''non-central'' in the sense of betweenness centrality. To take into account the difference in VK population of different cities, we considered both nominal and scaled values (by the number of users in the respective city) of betweenness centrality for each node in the city scale network. Overall, Table 1 shows that nominal and scaled betweenness centralities of most cities have remained relatively stable between 2015 and 2016 snapshots. A closer look at the obtained results reveals a combination of expected and surprising observations. On one hand, as one would expect, the largest cities in the post-Soviet space in terms of both actual and VK population (Moscow, Kyiv, and St. Petersburg) are in the top three in terms of betweenness centrality values, since these cities have so many VK users that larger fractions of shortest paths do pass through these cities. For example, approximately 16% of shortest paths in the city scale VK network go through Moscow, and 13% and 9% go through Kyiv, UA and St. Petersburg, respectively. However, when we scaled the respective values by VK populations of the cities, we observed that almost all of the cities with the highest values of betweenness centrality per user are located in Western Ukraine. This is interesting and surprising due to both the well-defined geographic pattern of these cities and the fact that Ukraine as a whole forms the largest and relatively ''tightly-knit'' modularity-based community

(with conductance value on a lower end of the spectrum as discussed above). Lastly, a surprising observation is that Yakutsk, which is located in a rather remote region of Russia, also has a high value of betweenness centrality per user.

Further, for the city scale network with aggregated edges (with the weight of each edge proportional to the volume of connection between the respective cities), we considered the subnetwork containing the largest cities in the post-Soviet space. Most of these cities are in Russia, Ukraine, Kazakhstan, and Belarus, where we observed a clear modular structure with higher volumes of connections between cities within the same country than between those in different countries. As it can be seen in the heatmap in Fig. 11, consistently with the aforementioned results on modularity-based communities, the separation of cities by country is the most pronounced for Ukraine, Kazakhstan, and Belarus, whereas the separation is not as strong but still visible for Russian cities. As one would expect, most of the squares on the diagonal are in the red/orange spectrum, which means that the highest volumes of connections are observed between users within the same city. This effect is the least pronounced in the case of Moscow due to the aforementioned considerations that many people in Moscow have strong social media ties with friends and/or relatives in other regions. More interestingly, we observed a clear trend of higher volumes of connections corresponding to cities that are geographically close to each other; moreover, this phenomenon is not affected by the presence of borders between countries. Specifically, it appears that this trend is the most visible in the cases where a pair of cities are within driving distance (that is, under 300-400 kilometers apart), which likely helps establishing social ties between people in those cities. For instance, there is a substantially higher volume of connections between Aktobe (Kazakhstan) and Orenburg (Russia) compared to any other Russia-Kazakhstan pair of cities. These cities are across the border but only 271 kilometers apart. Therefore, although inter-country borders do play a significant role in shaping the modular structure of the VK city scale network, the geographic proximity of cities (regardless of borders) is also an important factor.

Finally, the VK country scale network allows us to take a "bird's eye" view on the social media landscape of the post-Soviet space by considering the volumes of connections between and within the major countries in the post-Soviet space. The largest post-Soviet countries, Russia, Ukraine, Belarus, and Kazakhstan, have a higher volume of connections within the respective country than with other countries. However, smaller countries, including Azerbaijan, Armenia, and Moldova, have a higher volume of connections to Russia than within themselves or to any other country. This observation tells us that Russia still plays a centerpiece role in the VK network on the country scale, especially with respect to smaller post-Soviet countries, which suggests that these countries' economic and social ties with Russia as the largest country in the region are interrelated.

IV. METHODS

The data were collected solely for the purposes of basic scientific research by Social Media Analysis (SOMEA) group at the University of Jyväskylä, Finland (the group collects social media data for research purposes), using an application registered on VK.com working with a public API. A parallel crawler was utilized for the data collection, running on 15 servers. The collected data were not transferred to any other countries or institutions. The data were anonymized, that is, no personally identifiable information, such as names, user IDs, or photos of VK.com users, was collected or used in this study. The collected data and the presented results are not for commercial, advertisement, political, or any other use not related to scientific research. The data collection was in compliance with appropriate Russian, Finnish, and European Union legislations on personal data under provisions that the data were collected and processed solely for statistical/research purposes and anonymized. We used Google GeoCoding API (developers.google.com/maps/documentation/geocoding/) to fetch the coordinates and region information for about 250K cities populated by VK users. These were linked to the baseline graph. The final result of the retrieval process produced a 220GB TSV social graph. Using the C++ with boost, and Google performance tools (*goog-perftools*), the entire graph was loaded into memory with mapped files and processed. We have used two servers; the first server had 1TB of RAM and 64 cores; the second one had 192GB of RAM and 16 cores (as opposed to the study [7] that used Hadoop). This study used the residence information (city, country) indicated by the users on their accounts. Because this piece of information is not mandatory, not all users have submitted their location/country information. About 60% of users did provide the city/country information and we assume that the provided location information is correct. The results that deal with the geographical patterns within the VK network are based on the data for roughly 60% of users who provided their city/country information. In order to calculate shortest paths in the VK network, we have implemented the HyperANF algorithm [6]. The algorithm was modified, and instead of HyperLogLog counters, we used HyperLogLog++ [15]. The algorithm was implemented in C++ using OpenMP library for parallel processing. Matplotlib, amCharts, and Google Maps were used for drawing the respective figures.

REFERENCES

[1] N. F. Johnson *et al.*, "New online ecology of adversarial aggregates: Isis and beyond," *Science*, vol. 352, no. 6292, pp. 1459–1463, 2016. [Online]. Available: http://science.sciencemag.org/content/352/6292/1459

[2] N. Kinash, A. Tikhomirov, A. Trufanov, O. Berestneva, A. Boukhanovsky, and Z. Ashurova, "Analysis of large-scale networks using high performance technology (vkontakte case study)," in *Creativity in Intelligent Technologies and Data Science*. Cham, Switzerland: Springer, 2015, pp. 531–541.

[3] J. Leskovec and E. Horvitz, "Planetary-scale views on a large instant-messaging network," in *Proc. 17th Int. World Wide Web*, 2008, pp. 915–924.

[4] P. A. Longley, M. Adnan, and G. Lansley, "The geotemporal demographics of twitter usage," *Environ. Planning A, Economy Space*, vol. 47, no. 2, pp. 465–484, 2015.
[5] R. I. M. Dunbar, "The social brain hypothesis," *Evol. Anthropol.*, vol. 6, no. 5, pp. 178–190, 1998.
[6] P. Boldi, M. Rosa, and S. Vigna, "HyperANF: approximating the neighbourhood function of very large graphs on a budget," in *Proc. 20th Int. Conf. World Wide Web (WWW)*, New York, NY, USA, 2011, pp. 625–634, doi: 10.1145/1963405.1963493.
[7] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow. (2011). "The anatomy of the Facebook social graph." [Online]. Available: https://arxiv.org/abs/1111.4503
[8] B. Smriti, B. Moira, D. Carlos, F. I. Onur, and E. Sergey. *Three and a Half Degrees of Separation*. Facebook Research. Accessed: Jan. 10, 2018. [Online]. Available: https://research.fb.com/three-and-a-half-degrees-of-separation/
[9] M. E. J. Newman, "The structure and function of complex networks," *SIAM Rev.*, vol. 45, no. 2, pp. 167–256, 2003.
[10] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech., Theory Exp.*, vol. 2008, no. 10, p. P10008, 2008. [Online]. Available: http://iopscience.iop.org/article/10.1088/1742-5468/2008/10/P10008/meta
[11] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, "Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters," *CoRR*, Oct. 2008. [Online]. Available: http://arxiv.org/abs/0810.1355
[12] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *Proc. 19th Int. Conf. World Wide Web*, 2010, pp. 591–600.
[13] A. D. Broido and A. Clauset. (2018). "Scale-free networks are rare." [Online]. Available: https://arxiv.org/abs/1801.03400
[14] B. GonÃğalves, N. Perra, and A. Vespignani, "Modeling users' activity on twitter Networks: Validation of Dunbar's number," *PLoS ONE*, vol. 6, no. 8, p. e22656, 2011, doi: 10.1371/journal.pone.0022656.
[15] S. Heule, M. Nunkesser, and A. Hall, "Hyperloglog in practice: Algorithmic engineering of a state of the art cardinality estimation algorithm," in *Proc. EDBT Conf.*, Genoa, Italy, 2013, pp. 683–692.

ALEXANDER SEMENOV received the M.Eng. degree from Saint Petersburg State Electrotechnical University, Saint Petersburg, Russia, in 2008, and the Ph.D. dergee from the University of Jyväskylä, Finland, in 2013. During his doctoral studies, he also attended ITMO University, Russia. He was a Visiting Scholar with SUNY Buffalo, Buffalo, NY, USA, the University of Memphis, Memphis, TN, USA, and the University of Florida, Gainesville, FL, USA. He is currently a Post-Doctoral researcher with the Faculty of Information Technology, University of Jyväskylä. He has co-authored approximately 40 peer-reviewed publications. His research interests are in the analysis of large datasets, efficient algorithms, and machine learning. He was a recipient of research grants from TEKES, Finland, and EOARD, USA.

ALEXANDER V. MANTZARIS received the M.Sc. and Ph.D. degrees from The University of Edinburgh, in 2006 and 2011, respectively. He is currently an Assistant Professor with the Statistics Department. His research career began with investigating Bayesian methodologies that could improve the prediction accuracy in detecting horizontal gene transfer in bacteria DNA sequence alignments (2006–2010). The first research post he held involved applying machine learning techniques in a project devoted to studying the sonar behavior in bats to uncover the mechanisms in which they utilize it so effectively. His research interests include the areas of network science, big data, ad hoc network analysis, data analytics, and social behavior online.

ALEXANDER NIKOLAEV received the Ph.D. degree in industrial engineering from the University of Illinois, in 2008. His dissertation adviser was Dr. S. Jacobson. He is currently an Associate Professor and the Director of the Social Optimization Laboratory, Department of Industrial and Systems Engineering, University at Buffalo. His research group employs the techniques from both applied operations research and computer science. He has authored about 40 journals and about 10 peer-reviewed conference papers, with a citation count of around 300. His interests and expertise are in social network analysis, resource allocation under uncertainty, causal inference, healthcare, and educational data mining. He received three National Science Foundation Awards.

ALEXANDER VEREMYEV received the B.S. degree in mathematics and applied mathematics from Moscow State University, Moscow, Russia, in 2002, and the Ph.D. degree in industrial and systems engineering from the University of Florida, Gainesville, FL, USA, in 2011.

From 2011 to 2012, he was an Adjunct Assistant Professor with the Industrial and Systems Engineering Department, University of Florida. From 2012 to 2014, he was an NRC Research Associate with the Air Force Research Laboratory, Munitions Directorate, Eglin AFB, FL, USA. From 2014 to 2017, he was a Research Assistant Scientist with the Industrial and Systems Engineering Department, University of Florida. Since 2017, he has been a Research Assistant Professor with the Industrial Engineering and Management Systems Department, University of Central Florida, Orlando, FL, USA. His research interests include mathematical modeling, development, and analysis of efficient methods for solving various combinatorial optimization problems in complex networks.

Dr. Veremyev was a recipient of the National Research Council Research Associateship Award in 2012.

JARI VEIJALAINEN received the B.Sc. degree in mathematics and the M.Sc. degree in computer science from the University of Helsinki, Finland, in 1978 and 1983, respectively, and the Dr.Ing. degree in computer science from the Technical University of Berlin, Germany, in 1989. Since 1996, he has been a Full Professor with the Faculty of Information Technology, University of Jyväskylä. He was nominated to a Permanent Full Professor in software engineering in 1999. He was elected to become the head of the department in 1999, a position held by him also in 1997. From 2004 to 2005, he was heading the Information Technology Research Institute, University of Jyväskylä. He has been a Docent in computer science with the University of Helsinki, since 1996. His previous positions include permanent and visiting teaching and research positions at the University of Helsinki from 1980 to 1985, the Hahn-Meitner-Institut Berlin GmbH, Berlin, Germany, from 1985 to 1987, GMD-FOKUS, Berlin, in 1988, VTT, Espoo, Finland, from 1989 to 1997, GMD-IPSI, Darmstadt, Germany, in 1992, the Fraunhofer-Institut FIT, Sankt Augustin, Germany, from 2000 to 2001, GITI, Waseda University, Tokyo, Japan, in 2003, and the Hasso-Plattner-Institut, University of Potsdam, Germany, from 2005 to 2006.

EDUARDO L. PASILIAO, JR., received the B.S. degree in mechanical engineering from Columbia University, New York, NY, USA, in 1992, and the M.S. degree in coastal and oceanographic engineering and the Ph.D. degree in industrial and systems engineering from the University of Florida, Gainesville, FL, USA, in 1995 and 2003, respectively.

He is currently a Research Engineer with the Air Force Research Laboratory. He has co-authored over 100 peer-reviewed journal and conference publications. His research interests are in the areas of graph theory, combinatorial optimization, discrete mathematics, network science, and machine learning. He has been a member of the American Institute for Aeronautics and Astronautics and the Institute for Operations Research and Management Sciences.

VLADIMIR BOGINSKI received the B.S. degree in applied mathematics from the Moscow Institute of Physics and Technology, Russia, in 2000, and the M.S. degree and the Ph.D. degree in industrial and systems engineering from the University of Florida, Gainesville, FL, USA, in 2003 and 2005, respectively. He was an Assistant Professor of industrial and systems engineering with the University of Florida, Gainesville, FL, USA, and Florida State University, Tallahassee, FL, USA. Since 2015, he has been an Associate Professor of industrial engineering and management systems with the University of Central Florida, Orlando, FL, USA.

He has co-authored over 50 publications, including over 30 refereed journal articles. His research interests primarily focus on using mathematical modeling and optimization techniques in the emerging interdisciplinary area of network science. He is a member of the Institute for Operations Research and Management Sciences (INFORMS). Since 2016, he has been a member of council at the INFORMS Telecommunications and Network Analytics Section. He has served as the Vice-Chair for Networks at the INFORMS Optimization Society from 2013 to 2015. In 2018, he was selected by the U.S. National Academy of Engineering as a participant for the Japan–America Frontiers of Engineering Symposium, which brought together 60 of the most promising early career engineers from Japan and USA.