NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Olessia Y. Koltsova, Sergei N. Koltcov

WHEN LOCATION DOES NOT MATTER: MEMBERSHIP AND NETWORKING IN ONLINE COMMUNITIES OF SOFTWARE DEVELOPERS

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: SOCIOLOGY
WP BRP 57/SOC/2014

Olessia Y. Koltsova[1], *Sergei N. Koltcov*[2]

WHEN LOCATION DOES NOT MATTER: MEMBERSHIP AND NETWORKING IN ONLINE COMMUNITIES OF SOFTWARE DEVELOPERS[3]

Following the discussion on the role of Internet in the formation of ties across space, this paper seeks to supplement recent findings on location-dependent preferential attachment online. For this purpose, instead of looking at egonetworks, we look at networks of online communities specifically aimed at development of location-independent ties. The paper focuses on professional communities of software developers. The data are obtained automatically from the VKontakte social networking site. Evidence suggests that membership, friendship, commenting and liking ties are overwhelmingly cross-city

JEL classification: Z19.

Keywords: Internet, SNS, professional communities, cross-distance ties, networks, SNA.

[1] National Research University Higher School of Economics. Laboratory for Internet Studies. Director; E-mail: ekoltsova@hse.ru

[2] National Research University Higher School of Economics. Laboratory for Internet Studies. Deputy Director. E-mail: skoltsov@hse.ru

[3] The research is conducted in 2013-2014 with the support from the HSE Basic Research Program grant "Social & political processes on the Internet".

Introduction

Do online communities really extend possibilities of offline groups by permitting them to cross space, time or by decreasing the cost of communication? The empirical studies into these issues are still in their initial stage, and this paper seeks to address the question of space-dependence in professional communities of certain type. The concept of community itself has been dramatically transformed by the advent of the Internet, this notion being torn away from its attribute of local embeddedness. A new concept of space-independent virtual communities has emerged [Rheingold 2000], although within this notion multiple definitions are also possible. While traditionally sociological definitions of community underscore such necessary features as self-nomination, common identity and common goals [Rheingold 2000; Porter 2006], more empiricist definitions are based on the existence of real common activity or interaction [Smith et al. 2010].

Regardless of these variations, early reflections on the role of online communities and, more broadly, the Internet for societal bonds have expressed hopes that cyberspace may help people develop both common identities and interactions across space [Cairncross 1997]. Recent works, however, discover dependence of online ties on geographic distance or on the existence of off-line transport connections between places of users' residence [Goldenberg & Moshe 2009; Takhteyev et al 2011; Scellato et al 2010; Traud et al 2011]. Takhteyev and Scellato find that people tend to make friends with those who are closer, however, their conclusions are based on ego-networks. The latter might be thought of as naturally including those who are physically proximate. Traud et al deal with a specific sample of students of 100 universities where location is dormitory and members are by definition not geographically far. But certain kinds of communities, such as non-local social movements, mobile professional groups or groups of people with rarely occurring attributes, may be more inclined to develop ties across space and may be even specifically aimed at it. Thus, it makes full sense to look for space-crossing functions there where people are trying to make use of them.

For instance, McGee et al [2011], too, find that people tend to have their Twitter connections situated at the distance of approximately ten miles; then, however, they find an additional and in fact a much stronger peak at the distance of 1000 miles, which is especially pronounced for the relations of unreciprocated friendships and mentionings. They conclude that different types of relations and communication develop at different distances, and, specifically, that long-distance connections are used for information dissemination more than for reciprocated strong friendship ties. For our purposes, therefore, it may make sense, first, to look at different types of connections, not only at static friendship/followship, but also at more dynamic communication patterns, such as

commenting, liking or mentioning. Second, it may make sense to measure the inclination toward space-crossing connections not as a proportion of individual links, but as a proportion of links in a community that declares a goal to connect people independently of their location.

Such an attempt is made in this work, with same-city links being considered "local" and between-city links viewed as "space-crossing". The rest of the article is structured as follows. In the next section we reflect on types of social actors who could wish to connect across space and the reasons why they might want to do so, which leads us to the selection of software developers as a professional group for further analysis. Next, we describe our data, the stages of sampling and preliminary analysis of membership and friendship network structure of several selected groups, and justify the final selection of a single group for in-depth case study. In the following two sections we give a more elaborated analysis of the selected group's friendship network and of its posting, liking and commenting activity and networks. All analysis is done in terms of its relation to geography. The article ends with the conclusion section.

Who and why could develop ties across space

So what types of communities would seek to cross geographical space? Groups aimed at improving their neighborhoods or consumer groups centered around locally routed services would not be the candidates. However, worldwide social movements would be constrained only by language barriers, not by distance per se. For instance, Facebook groups have been critical in the spread of Occupy movement across the globe [Caren & Gaby 2011]. Likewise, in the past decade people or relatives of people with rarely occurring diseases have tended to form very successful support communities online [Lasker et al 2005]. One of the reasons is that diseases considered "rare" locally turn to be relatively wide-spread on a national scale: for instance, US National Institute of Health's Office of Rare Diseases recognizes as rare all diseases that affect fewer than 200,000 people in the United States, which is potentially a huge community [Lasker et al 2005]. One another candidate to form cross-distance communities would be professionals whose careers could somehow benefit from joining professional virtual worlds – either by providing publicity and job opportunities, or knowledge sharing and competencies gaining, or anything else. Recent research shows that professional communities have proliferated particularly with certain kinds of occupations, such as medical workers, educators and software developers / IT professionals, as well as students of respective professions [see e.g. Thompson et al 2008; Duncan-Howell 2010; Barcellini et al 2005].

This latter group has deserved special attention because it is developers who literally coordinate their collaborative software projects online, especially aimed at open source products. Thus for them online community is not only a place for talking about their work, but actually for doing their work together, but at a distance. *Collaboration*, thus, is one of motivations to join a cross-space professional community that we would like to list here, although it is not the only one. In fact, in their study of GitHub, a large developers' community, Wu et al [2014] have found only a weak correlation between followship and production activities which means that developers may connect to each other for reasons other than collaboration. Singer et al [2014] have established that software developers use Twitter to stay aware of industry changes, to follow specific projects and to promote their own ones, as well as for learning, and for building professional relationships and networking.

Awareness of changes is important for fast-developing professions, to which software engineers definitely belong. Moreover, centers of innovation or just of up-to-date knowledge may be geographically far, and this is a distinct motivation to develop ties across space. *Learning* is very close to awareness, still it presupposes concrete knowledge and skills acquisition rather that monitoring. While for such professionals like medical workers or others having a large portion of manual skills it is not easy to share them online and at a distance, for developers whose work is entirely within a computer it is much easier to share skills through computer.

Networking may be useful, as mentioned above, for developing collaborative projects, but also for finding new jobs, and here software developers may especially benefit not only because they can work at a distance, but also because this group is one of the most open for geographical and job mobility, the latter being well-studied [Huang 2012]. Therefore, visibility and self-promotion are quite important. Wickramasinghe and Weliwitigoda [2011] find out that the number of networks of which an individual is a member and the frequency of interaction between network members significantly predict job security and career advancement of software developers in Sri Lanka, so inclination to online networking might be expected from this professional group.

Finally, *identity development and maintenance*, as well as feeling of belongingness to the professional subcultures, may be also important for occupations with strong professional identity. Software developers are among those whose identity is very strong [Marks & Scholarios 2007], while it is known from other studies that online communities play an identity-building role for other professions [Hara & Hew 2007, Goose & Bennison 2008]. Since many developers work at homes, isolated from immediate communication with their colleagues, maintaining a strong identity almost inevitably demands computer-mediated forms of communication and promotes joining online communities, still this type of motivation has no geographical constrains. Basing on all these

reflections, we have concluded that software developers could be one of professional groups most inclined to develop connections across space.

Data and sampling

Many communities, especially grassroots ones, often use social networking sites (SNS) as their platforms because the threshold of entry to the virtual space through an SNS is very low – actually, it is free. Another feature of SNSs is that their group pages have a unified format and therefore are very convenient for comparison; altogether, this could make SNS groups very attractive for researchers. However, not much comparative research is done on SNS groups and quite little is known about their general features, due to the closed character of the World's main SNS, Facebook. It prevents efficient data collection, unless they are received from Facebook itself like in [Traud et al 2011; Backstrom et al 2012]. To our knowledge, all Facebook research is devoted to egonetworks of friendships between accounts that are not self-selected into community pages [Bryant & Marmo 2012; Catanese et al 2010]. However, FB's Russian equivalent, Vkontakte (VK), that is much more popular in the Russian-language space than the FB itself, allows rich data collection from its open individual and community pages. Therefore, an explorative study of community page's features has been one of the goals of this research, too.

There is no perfectly reliable general VK statistics (which is necessary as the baseline for this research). VK official catalogue lists about 250 million of individual accounts, and a list of about 75 million group pages may be extracted from its server. A large proportion of them are, however, deactivated or just inactive. The most comprehensive publicly available statistical overview of individual users was done by a free-lance expert Ilya Demyanenko in 2011 and published at the blog platform for software developers, habrahabr.ru [Демьяненко 2011]. These data are used here for background comparisons, albeit they should be treated with caution. Our own research of a random sample of 150,000 groups shows a highly skewed power law distribution of group sizes and numbers of messages, which suggests that sampling should be made from the top rather from the middle.

VKontakte contains 3,965 groups with the word "programmist" (software developer) in the title, as of October 3, 2014, the number of members and posts being exponentially distributed, like in the random sample. At the first stage we briefly examined all groups with over than 10,000 members (see Table 1). Like in Facebook, communities in VK are divided into "pages" and "groups", the latter being either open or closed (that is, available to everybody or to the approved users only). Each open VK group contains the following fields: general info, picture, administrators' contacts

and hyperlinks that the group finds the most important; the list of members; the "wall" – a message board with dendroid structure that allows posting, commenting on posts, liking posts and liking comments to all VK visitors; space for launching discussion threads with a more linear structure (posts and likes), and space for uploading photos, videos, audios and documents. VK pages have subscribers instead of members, they have no space for discussion threads and do not allow posting on the wall, although commenting or liking may be allowed; any other content contribution is not permitted. Groups, however, can also restrict access to any of their fields or functions. Pages and highly restricted groups resemble regular media more than communities; indeed, the majority of popular pages in VK are moderated by a small number of media professionals or even companies whose income comes from advertising in those pages. Thus, in our sample three of 13 communities are owned by the same company, tproger.ru, two of them are moderated by the same person, and all of them are pages. Therefore, we excluded all pages from the analysis. Also, *JavaRush* was excluded because it is a VK extension of a Java teaching center and, like a page, permits no posting, and *Programmer of Delphi, Pascal* was excluded because it is virtually inactive.

Tab. 1. Top programming-related groups in VKontakte

	Group title	Type	Members/ subscribers	Messages on wall	Number of discussions
1	Typical programmer	Page	281 452	7 675	0
2	ITmozg — programmers, designers	Page	102 146	4 056	0
3	JavaRush | I want to become a Java programmer	Open group	81 468	1	0
4	Girl programmer	Page	32 189	3 428	0
5	Programmers	Open group	31 038	1	8 174
6	Programmer's son	Page	28 832	3 339	0
7	Programmer's library	Page	25 663	637	0
8	Programmers C/C++	Open Group	19 291	37	2 558
9	All-Russia learning marathon "Programmer"	Page	16 515	34	0

10	Programmer (former C#)	Open group	15 430	1 626	39
11	Atypical programmer	Page	11 379	10	0
12	Programmer's reference book	Open group	11 238	795	331
13	Programmer of Delphi, Pascal	Open group	11 138	158	1

At the second stage we downloaded all members and friendship links between them in the remaining four groups – *Programmers, Programmers C/C++*, *Programmer* (former C#) and *Programmer's reference book*; further on they are termed *Progs*, *C/C++*, *C#* and *Refbook* for convenience. Progs and *C/C++* state Russia as their location, *Refbook* states Yekaterinburg, a city in the Urals, Russia, and *C#* states the city of Odessa in Ukraine. All data in this research was downloaded with VKMiner, the software developed in the Laboratory for Internet Studies for this and other projects.

Before analyzing geography of these groups, it is necessary to comment on deactivated users observed in them: most often, the VK administration suspends accounts suspected of spam or other undesirable activity. Many groups struggling for SEO buy services of creating fake accounts and adding them to groups to increase their population and to get into the top of the search results. Thus, the share of deactivated users is a crude indicator of the degree to which a particular group has been trying to commercialize. Although gender and friend lists of users are available after their suspension, in Table 2 we calculate most measures based on the list of non-deactivated accounts. Also, all four groups do not differ much from the general VK in terms of age distribution, and all are male-dominated, albeit to a varying degree.

Geography of friendship in the four groups

Coming to the spatial analysis, it should be noted that space and ability/inclination to connect across space can be conceptualized in a number of different ways. For instance, they may be measured as exact distance between users. However, first, we do not know the distance between same-city VK users. Second, and most important: exact distance between cities in such spaces as Russia and post-Soviet world seems not the most relevant, since even "small" distances are most often too large to commute, transport connection is poor and expensive, and most kinds of "offline" social life tend to be clustered within cities or villages. The latter, however, contain very few Internet users. Thus dependence on distance, if any, would most probably obscure dependence on the belonging to the same city. That is why this belonging was selected as a proxy for the geographical "sameness".

Tab. 2. Basic features of the four developers' groups.

	Progs	C/C++	C#	Refbook
users total	31 003	19 333	15 452	11 238
share of deactivated	25.74	10.31	34.15	23.07
share of females, among non-deactivated	17.25	17.12	32.93	37.86
share of Russia, among having country	67.15	63.29	61.83	60.62
number of cities	2 034	1 715	1 447	1 362
ratio non-deactivated users to cities	11	10	7	6
share of Moscow, among having city	14.01	13.57	15.38	14.81
share of St.Petersburg, among having city	10.74	9.17	5.96	5.67
edges total	50 872	17 828	50 915	21 713
Density	0.00011	0.00010	0.00043	0.00034
share of isolates, among non-deactivated	56.71	56.18	49.67	64.70
mean degree, among non-deactivated	1.90	1.74	6.97	4.19
all possible friendships in sample	50 265 351	20 247 066	9 841 266	3 512 575
all possible same-city friendships in sample	1 891 904	749 936	567 630	207 801
all existing friendships in sample	20 149	11 899	27 704	15 475
all existing same-city friendships in sample	5 484	4236	1 729	1 998
total density in sample	0.00040	0.00059	0.00282	0.00441
overal within-city density in sample	0.00290	0.00565	0.00305	0.00961
between-city density in sample	0.00030	0.00039	0.00280	0.00408
posssible same-city to possible non-same-city ratio	0.03911	0.03704	0.06121	0.06288
existing same-city to existing non-same-city ratio	0.37395	0.35600	0.06656	0.14825

share of possible same-city links among all possible links	3.76	3.70	5.77	5.92
share of all existing same-city links among all existing links	27.22	35.60	6.24	12.91

For the four selected groups, basic descriptive statistics and a number of network measures are presented in Table 2. In these four groups, 86-88% of non-deactivated users indicate their cities, although the subsets of linked pairs where both vertices have cities embrace only 31-44% of all non-deactivated accounts. Unfortunately, we have not found an appropriate way of imputation of such vertex attributes as city: this variable can take dozens of thousands of values, most of which are not present in the studied set. Probability of them occurring in it is far from random, but largely unknown – in fact, had it been known, much of this research would have been redundant. Therefore, we had to delete all pairs with missing cities and bear with the limitations of such approach.

The mere city count already shows that users are scattered across a large number of cities (see Table 3). Their distribution is highly skewed, but reflects the general VK city distribution [Demyanenko 2011], to the extent to which it is known. Moscow is slightly overrepresented in all four groups, while St.Petersburg is visibly overrepresented in *Progs* and *C/C++*. Kiev, Ukraine's capital, and Misnk, Belorussian capital, hold the third and the fourth positions respectively in all groups, while the cities that follow present a mixture of Russian and Ukrainian places in the proportion close the general VK proportion. Most cities in the long tail are represented by a single user. Thus, if belonging to the same group is considered a tie, software-interested users clearly develop these ties irrespectively of their locations.

A stricter approach, however, is to look at friendship relations within these groups. Again, the fraction of same-city (or within-city) friendships is smaller than that of between-city friendships, but varies greatly: from a quarter and a third in *Progs* and *C/C++* respectively to only 6% in *C#*, the latter being not much different from the respective "possible share". That is, if the *C#* friendship graph was full, same-city friendships would constitute 5.77% of all friendships, while in the observed graph they constitute 6.24%. Although it is impossible to generalize from only four groups, it is worth noting that the two groups with higher proportion of same-city friendships are, first, "biased" in favor of St.Petersburg, and second, unlike the other two, limit their wall activity channeling it to the topical discussions section.

Tab. 3. Developers' groups distribution by city

VKontakte in general (2011)		Progs		C/C++		C#		RefBook	
City	Share	City	Share	City	Share	City	share	City	Share
Moscow	10.42	Moscow	14.02	Moscow	13.57	Moscow	15.42	Moscow	14.81
St.Petersburg	5.80	St Petersburg	10.74	St Petersburg	9.17	St.Petersburg	6.03	St Petersburg	5.67
Kiev	2.79	Kiev	4.08	Kiev	4.42	Kiev	5.4	Kiev	5.20
Minsk	1.73	Minsk	1.90	Minsk	2.49	Minsk	2.05	Minsk	1.65
Yekaterinburg	1.51	Kharkiv	1.79	Kharkiv	2.45	Yekaterinburg	1.44	Dnipropetrovsk	1.54
Novosibirsk	1.36	Novosibirsk	1.55	Novosibirsk	1.86	Dnipropetrovsk	1.41	Kazan	1.36
Kharkiv	1.23	Kazan	1.51	Kazan	1.64	Odessa	1.41	Krasnodar	1.36
Samara	1.14	Nizhny Novgorod	1.38	Nizhny Novgorod	1.64	Novosibirsk	1.37	Kharkiv	1.25
Nizhny Novgorod	1.10	Yekaterinburg	1.31	Odessa	1.36	Kharkiv	1.36	Donetsk	1.23
Omsk	1.03	Dnipropetrovsk	1.30	Dnipropetrovsk	1.32	Kazan	1.27	Odessa	1.19
		Odessa	1.16	Lviv	1.24	Donetsk	1.18	Yekaterinburg	1.19
		Rostov-on-Don	1.08	Donetsk	1.23	Lviv	1.17	Almaty	1.09
		Samara	1.05	Yekaterinburg	1.16	Chelyabinsk	1.1	Chelyabinsk	1.07
		Krasnodar	1.04	Krasnoyarsk	1.10	Nizhny Novgorod	0.93	Lviv	1.07
		Ufa	1.01	Ufa	1.04	Samara	0.92	Ufa	1.02

Similar situation is with the same-city to between-city links ratio: it is in favor of between-city links, but in the two St.Petersburg dominated groups it is much higher than the "possible ratio", while in *C#* the observed ratio does not differ much from the ratio possible in the full graph. The tendency repeats itself in the relationship of within-city density and between-city density, with *C#* having the most similar values of these two densities. It, therefore, may be concluded that members of software-related groups befriend each other across cities much more actively than within their cities, however, in some groups city "sameness" would increase their willingness to befriend each other. *C#*, where city homophily is minimal, was selected for further analysis of its structure and content, including commenting and liking networks.

C# group and its networks

C# is the only group in the sample that indicates its "location" outside Russia (in Ukraine). However, location of VK groups may be conventional – thus, the *C#* moderator is situated in Minsk, Belarus, and the distribution of *C#* users over countries does not differ from other groups (Table 4) (90% of active accounts indicate their countries). Nearly all communication in the *C#*

group is in Russian, with only eight posts being in Ukrainian and one in English. The Ukrainian crisis seemed to have little effect on the group: there was one crisis-related announcement and one video reposted by the administrator. The fact of reposting it to the software group aroused sharp criticism from one of the members and a discussion between him and the administrator in the thread of comments that, however, faded away very fast.

Tab. 4. Developers' groups distribution by country

Progs		
country	Users	%
Russia	14 165	67.15
Ukraine	4 374	20.74
Belarus	884	4.19
Kazakhstan	468	2.22
USA	283	1.34
Germany	99	0.47
UK	45	0.21
Azerbaijan	42	0.20
Israel	42	0.20
Moldova	37	0.18
Italy	35	0.17
Kyrgyzstan	32	0.15
Poland	30	0.14
Japan	28	0.13
Canada	28	0.13

C/C++		
country	users	%
Russia	9 940	63.29
Ukraine	3 670	23.37
Belarus	735	4.68
Kazakhstan	355	2.26
USA	213	1.36
Germany	85	0.54
Armenia	53	0.34
Israel	44	0.28
UK	34	0.22
Moldova	33	0.21
Tajikistan	33	0.21
Uzbekistan	31	0.20
Italy	29	0.18
Kyrgyzstan	27	0.17
Azerbaijan	23	0.15

C#		
Country	users	%
Russia	5 725	61.83
Ukraine	2 265	24.46
Belarus	582	6.29
Kazakhstan	193	2.08
USA	177	1.91
Germany	36	0.39
Moldova	32	0.35
Azerbaijan	27	0.29
Armenia	20	0.22
UK	12	0.13
Spain	12	0.13
Latvia	12	0.13
Japan	11	0.12
Israel	11	0.12
Uzbekistan	11	0.12

RefBook		
country	users	%
Russia	4 804	60.62
Ukraine	1 971	24.87
Belarus	434	5.48
Kazakhstan	318	4.01
USA	116	1.46
Moldova	45	0.57
Armenia	27	0.34
Azerbaijan	19	0.24
Kyrgyzstan	15	0.19
Germany	14	0.18
UK	10	0.13
Spain	10	0.13
Georgia	10	0.13
Latvia	9	0.11
Italy	8	0.10

By the time of data collection, the C# group contained 15 451 members, 895 posts, 1 392 comments and 1 595 likes on the wall. However, of 1 951 users who contributed to the wall only 806 (41.3%) were group members. Given that most activity in this group takes place on the wall, with discussion threads containing only 260 messages, this poses a question on the boundaries of the community. The situation is not unique: most VK groups studied also consist of a silent majority and an active core that is not entirely situated within the group. Another challenge posed to the issue of group boundaries is the abundance of deactivated accounts in VK: as mentioned above, in the studied

group a third had been deactivated by the time of data collection. In addition, 170 of 1145 (14.8%) guest contributors to the wall had been also dead. Luckily, they have created only a little more than a hundred posts and comments, however, dead contributors account for almost one third of likes and dead group members account for approximately one third of friendships, with their mean degree being just a little lower than that of the "living" users.

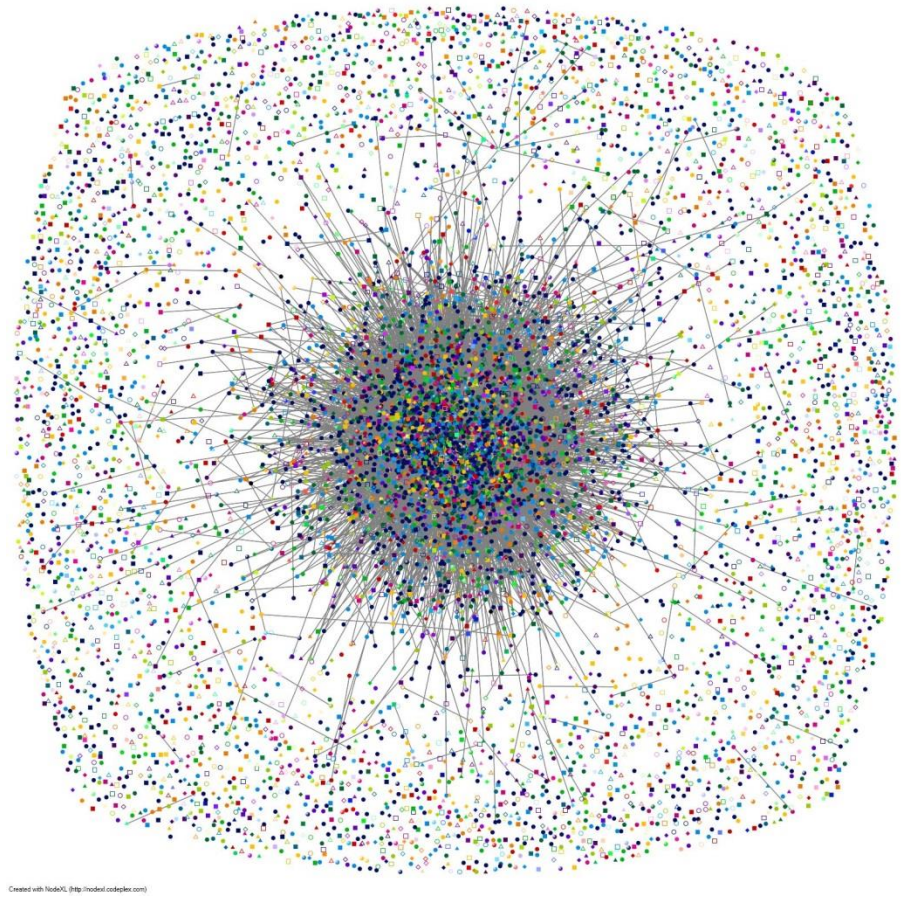

Fig. 1. C# friendship network. Combinations of color and shape denote cities.

With deactivated accounts being filtered out, the friendship network structure has been studied and it revealed a structure found in most other studied VK groups, and in fact in many other online graphs [Lescovets et al 2008]. Namely, it contains a single giant component with no clear underlying structure (modularity =0.29), a small number of very little components and a large periphery of isolates that constitute the half of the active accounts (Figure 1). Commenting and liking networks show a different structure (Figures 6 and 7) since most contributors leave only one comment or like, these networks are collections of stars loosely connected by a minority of those who have chosen to leave 2-3 instances of feedback. The centralization around local stars and loose connection between them is especially visible in the liking network with its modularity of 0.85 against 0.4 in the commenting network. Half of the posts receive no comments, and 80% of them are never liked, so the networks are really small. Contributing group members are not more, and

even a little less central in the networks of friendship than silent members, which additionally suggests that communicative activity, on the one hand, and networking and self-identification with the group, on the other, are disconnected processes.

Friendship network in C# group

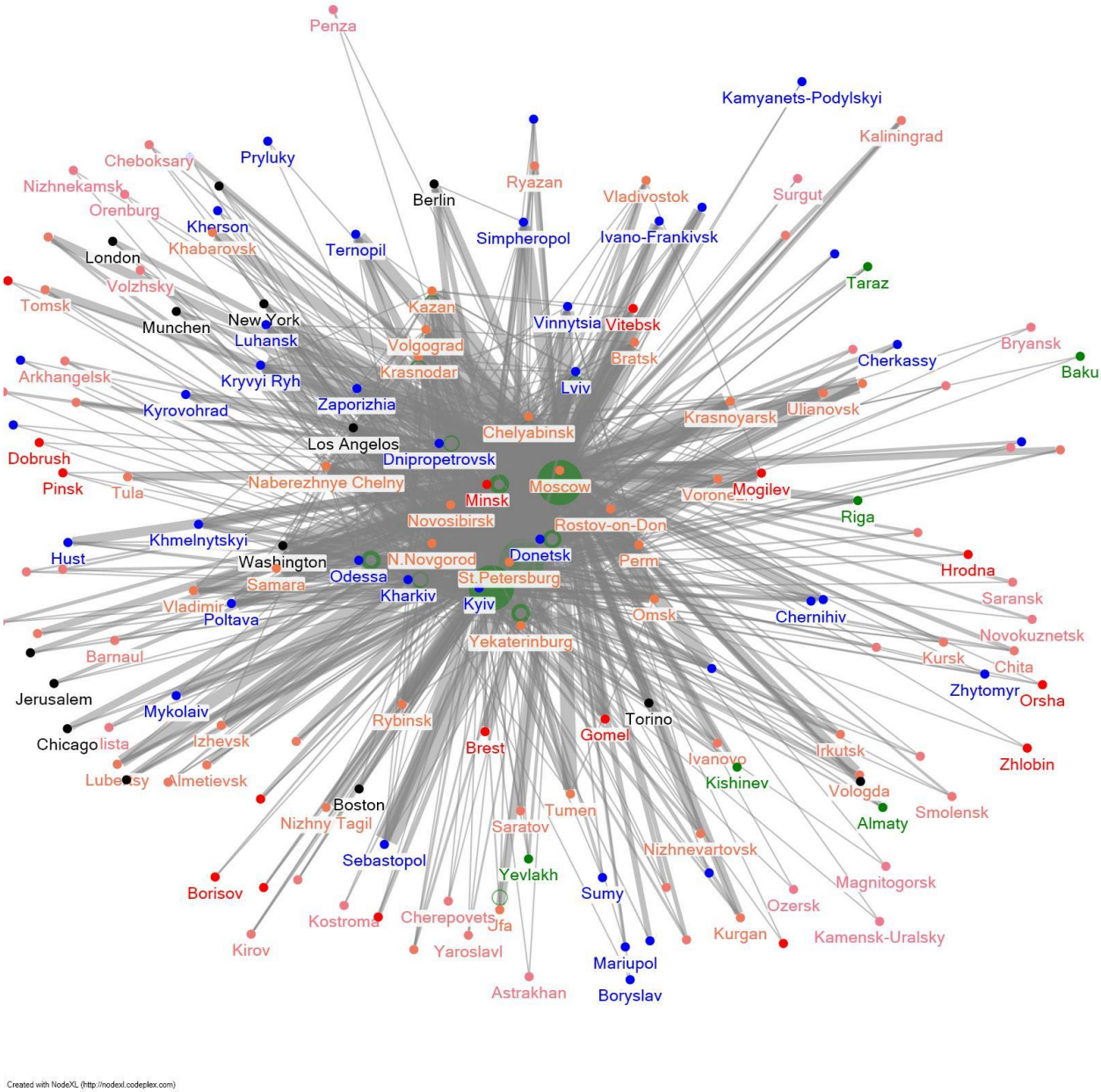

Fig. 2. Cities with 5+ friendship ties to at least one another city. Vertices: groups of users from the same city. Colors: orange – Russia, Blue – Ukraine, Red – Belarus, Green – other Post-Soviet, Black – "far foreign" (other non-Russian).

Fig.1 shows overall core-periphery ratio in the C# group, still it fails to show how users from different cities are connected or disconnected. In Fig. 2 we collapse all users from a given city into a single vertex labeled with the name of the corresponding city. Friendship ties of all same-city users

become the edges of such new city-level vertex. Thus multiple links between these new vertices are possible, but they are also merged and the graph is made weighted. For better visibility, we show only those cities that have at least one edge with weight = 5 or more. Friendships within cities thus become self-loops, they are shown in green, and it can easily be seen that they are very few (only 15 out of 157 possible on this graph, against 8 373 inter-city edges).

It may be argued that although users clearly prefer to connect beyond their own city, some pairs or sub-groups of cities may have preferential attachment to each other. However, modularity of the shown graph is 0.25, and the full graph that includes low-weight edges has even a lower modularity of 0.15. Instead of clustering, we again see a strong center-periphery differentiation. Three cities with highest degrees – Moscow, Kyev and St.Petersburg – clearly dominate in connecting to others (fig. 3a, 3b and 3c highlight their ties correspondingly). Other well-connected cities have their thickest ties with those three (fig. 3d, 3e and 3f show example cities from Russia, Ukraine and Belarus). Thus some cities serve as hubs more than others, however, the number of out-links is just a function of city size (correlation is over 0.9).

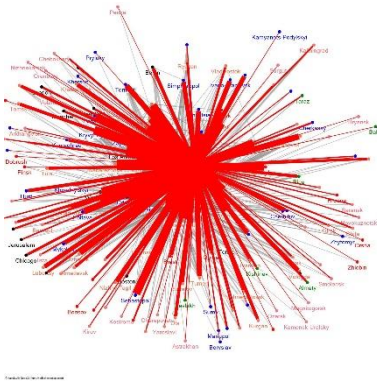

Fig. 3a. Moscow

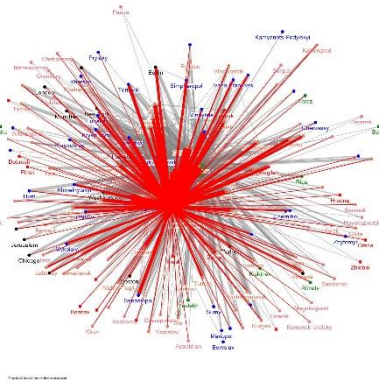

Fig. 2b. Kyiv

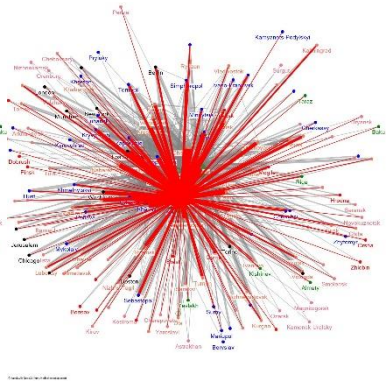

Fig. 3c. St.Petersburg

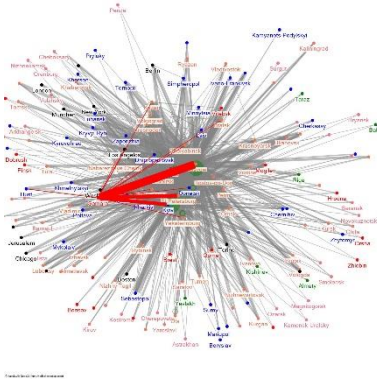

Fig. 3d. Samara (Russia)

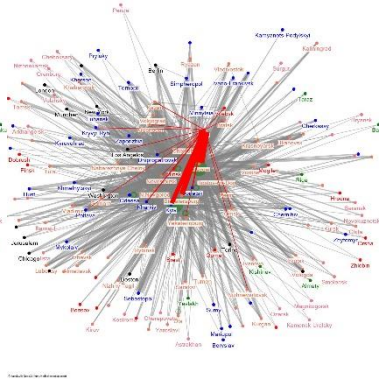

Fig. 2d. Lviv (Ukraine)

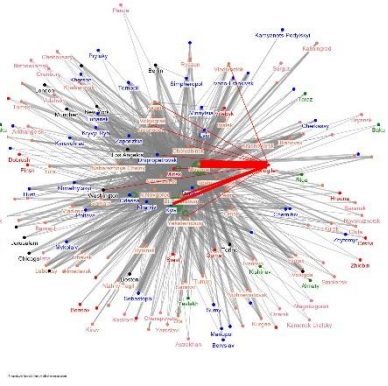

Fig. 3f. Mogilev (Belarus)

In the previous sections we also noted that the within-city density is not much different from between-city density in the friendship network of the C# group. However, after finding this we have

hypothesized that distribution of within-city densities may be uneven, and that a few cities in fact can form dense sub-communities, while the overall low value of modularity may conceal this fact.

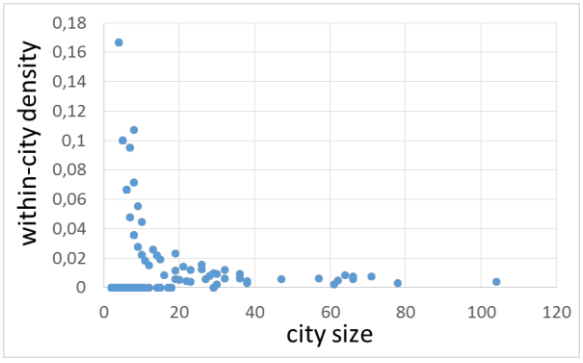

Fig. 4a. Relationship between city size and friendship network density, C# group, four largest cities not shown

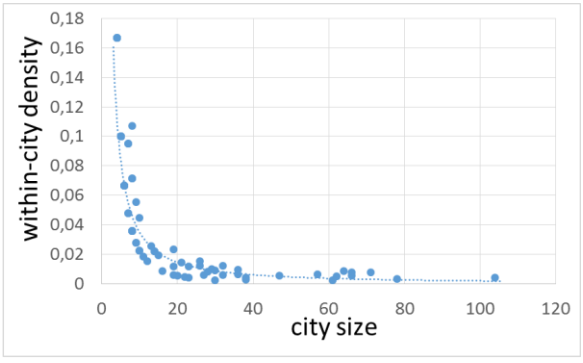

Fig. 4b. Relationship between city size and friendship network density, C# group, zero-density cities and four largest cities not shown

Figure 4a shows that the distribution of in-city densities is a mixture of two distributions. A large proportion of cities with less than 20 users demonstrates zero density; when they are filtered out (Fig. 4b), the remaining distribution is best approximated by a power-law function. That is, density is highly (and negatively) dependent of the city size, which happens just because it is calculated as the number of existing links divided by nearly squared group size. Thus the studied sample follows the universal trend, and there no outliers. This means that no unusually dense communities within any cities are observed.

We then hypothesized that, despite absence of unusually dense cities, there may be cities that are unusually disconnected from others. Fig. 5 plots cities in the space of in-city and out-city friendships. It shows that most cities have less than five in-group links, while quite many of them possess a few hundreds of outlinks. Even those cities that are below the trend line have in-to-out ratio less than one (except a few with 1-4 users). For comparison, *Progs* group contains only one city with the in-to-out ratio of more than one, Grodno, a town near the Belorussian-Polish border (again, except a few with 1-5 users), and not St.Petersburg which is overrepresented in this group. Therefore, both groups have nearly no outliers, but the slope of trend line for *C#* is steeper, which means that all cities relatively evenly contribute to its lower proportion of within-city links.

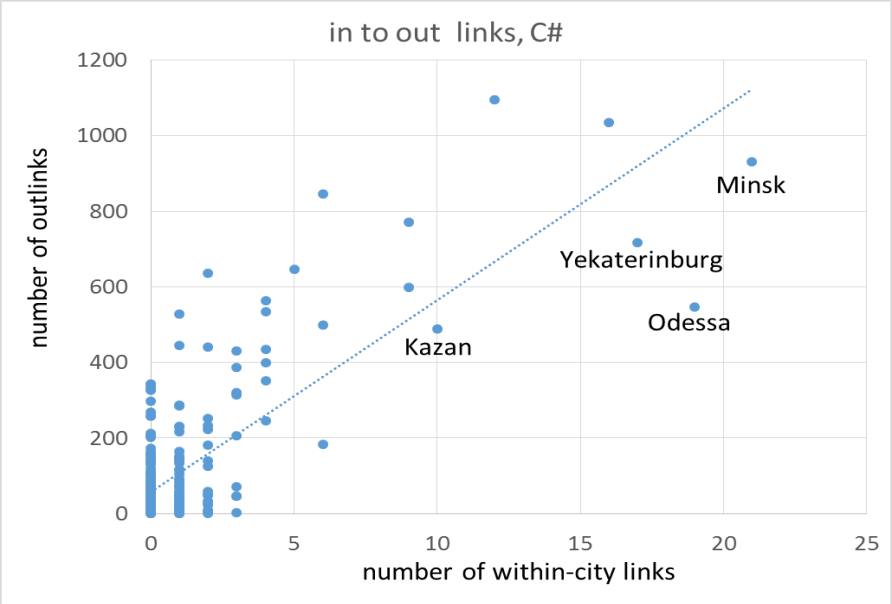

Fig. 5. In-city to out-city ratio distribution, C# group

Commenting and liking features in C# group

Within commenting and liking, several kinds of activity/feedback can be singled out: contribution through posts, comments, likes, and feedback received in comments to posts, likes to posts, and likes to comments. None of them correlates, except commenting and liking, which means that users may be divided into posters, on the one hand, and feedback givers, on the other. Guest contributors and females prefer posting to commenting with group members and males having the reverse ratio of posts and comments. Gender differences are more pronounced. Posts may be divided into the following categories with descending frequency: questions & requests for advice in software development, hardware choice and use, and other computer-related tasks, search for people to solve a task for free or for payment, job announcements, software developing service offers, search for existing software, and other announcements and jokes. Comments mostly offer solutions; therefore, while females and outsiders tend to seek for advice and help, male insiders tend to give feedback.

The whole space is male-dominated: among members, males prevail with ratio 2:1, among contributors – 3:1, while overall gender distribution in VK is nearly even. Males are also a little more central in the friendship network. Age analysis is hindered by incompleteness of these data: overall, less then one-fifth of VK accounts contain data on age; in our sample – 30% of all accounts and 44% of active accounts. Based on these data, it can be said that the age distribution in the sample is slightly different from that of VK in general: the VK distribution sharply peaks at the age of 24-25, being nearly symmetrical in the range 14-34 and receding into the long tail at older ages.

In the group, distribution in the range 15-23 is nearly flat and drops after that. Around a half of the group are users younger than 24, that is, of college and school age, which means they can not be professional software developers yet. Among contributors, this "24-" cohort constitutes 60%. Indeed, many requests for help list university tasks to be solved, and some even offer student help as paid service. It is worth noting that the two discussion-dominated groups prohibit these two kinds of posts except in the special discussion threads.

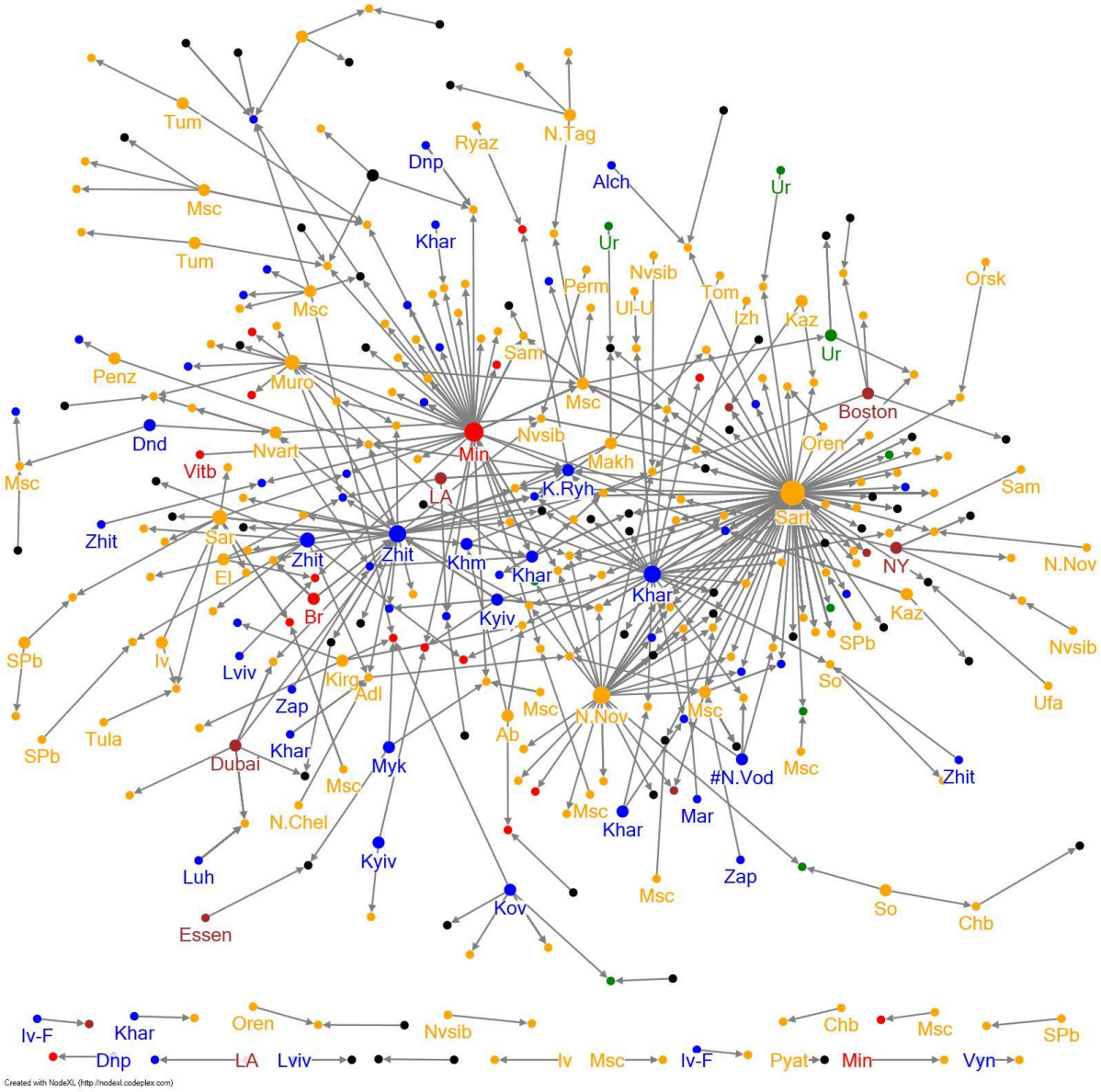

Fig. 6. Commenting, C# group. Vertices: individuals. Size: out-degree. Color: orange – Russia, blue – Ukraine, red – Belarus, green – other Post-Soviet, brown – "far foreign", black – unknown. Vertices with 0 in-degree not labeled. Arrows indicate direction of commenting. Self-comments filtered out.

The subset of the commenting network for spatial analysis embraced 71% of edges and 62% of vertices, due to the incompleteness of city data (Fig. 5). In-degrees in this network did not exceed 5, while top six commentators had out-degrees between 12 and 73 which means that the network was

dominated by activity hubs, not by popularity stars. Hubs were connected through multiple users who had got commented by several of those active commentators, but had not left a single comment themselves. All this corresponds to advice-giving network where a limited set of experts answers questions of occasional visitors.

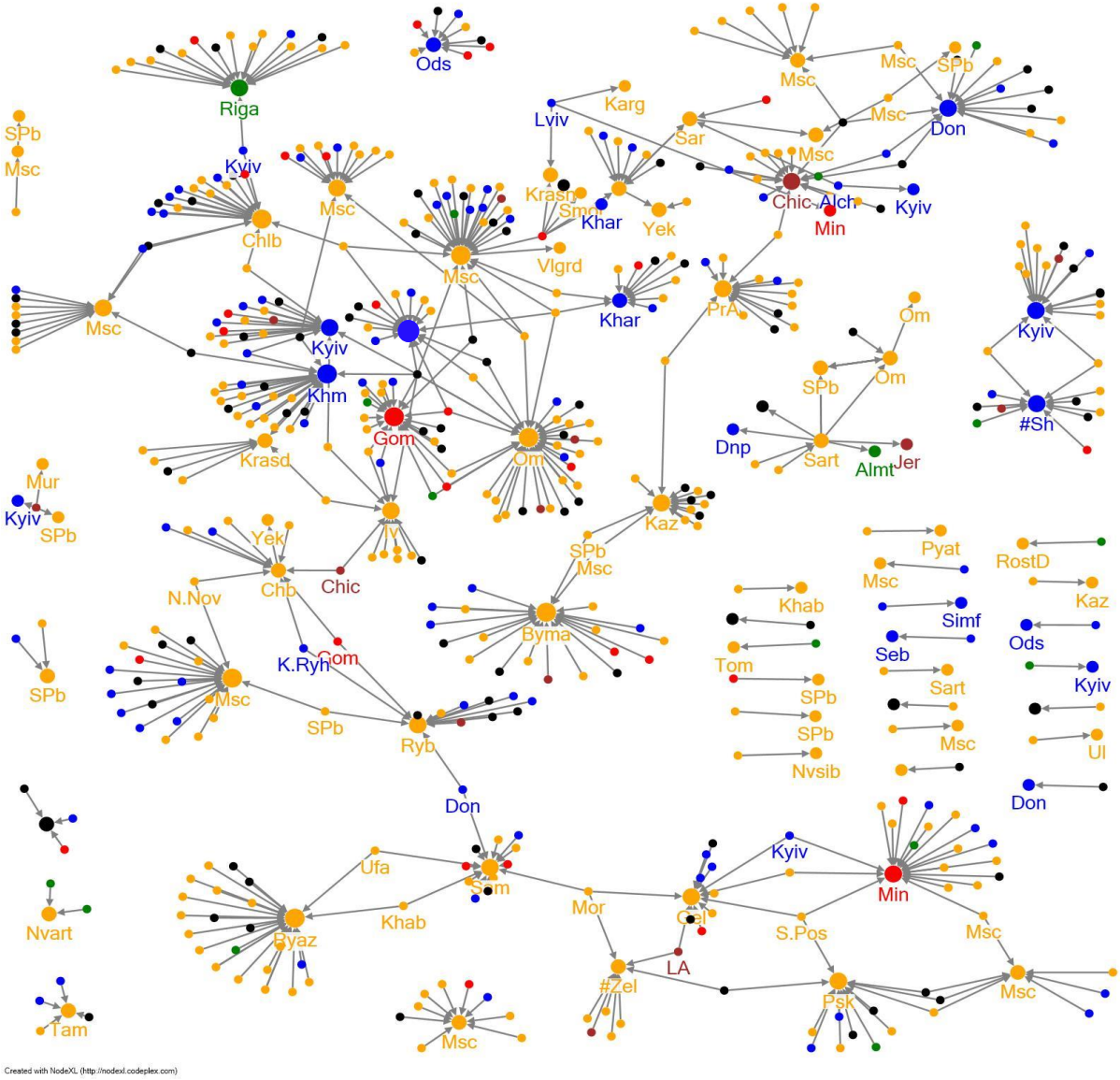

Fig. 7. Likes on posts, C# group. Vertices: individuals. Size: in-degree. Color: orange – Russia, blue – Ukraine, red – Belarus, green – other Post-Soviet, brown – "far foreign", black – unknown. Vertices with 0 in-degree not labeled. Arrows indicate direction of liking.

Geographical analysis of commenting and liking networks is a little more difficult than that of friendship because the notion of the maximal number of possible links is not applicable here: more than one edge between each pair of vertices may exist, and each of them is meaningful. Still, if the network that included all commentators and all commented posters was full, and each member could give only one comment, it would have about 2.7% of same-city links. The real proportion is

31%. However, nearly all same-city comments are self-comments – that is, replies of the poster to his/her commentators in the thread of comments to this post. When these self-comments are filtered out, only eleven same-city links remain, which constitutes 1.6% of the full network - actually lower than it could be expected even if each user could comment another user only once. It should be noted, however, that many posts have clear calls to connect through by private messaging, especially job announcements and requests for elaborated problem solving. Some job announcements state the city of the opening vacancy; therefore, it is unclear how much same-city communication happens in private messages. If it is large, then it may be concluded that the group is, among other things, used to connect those who are geographically close but do not know of each other.

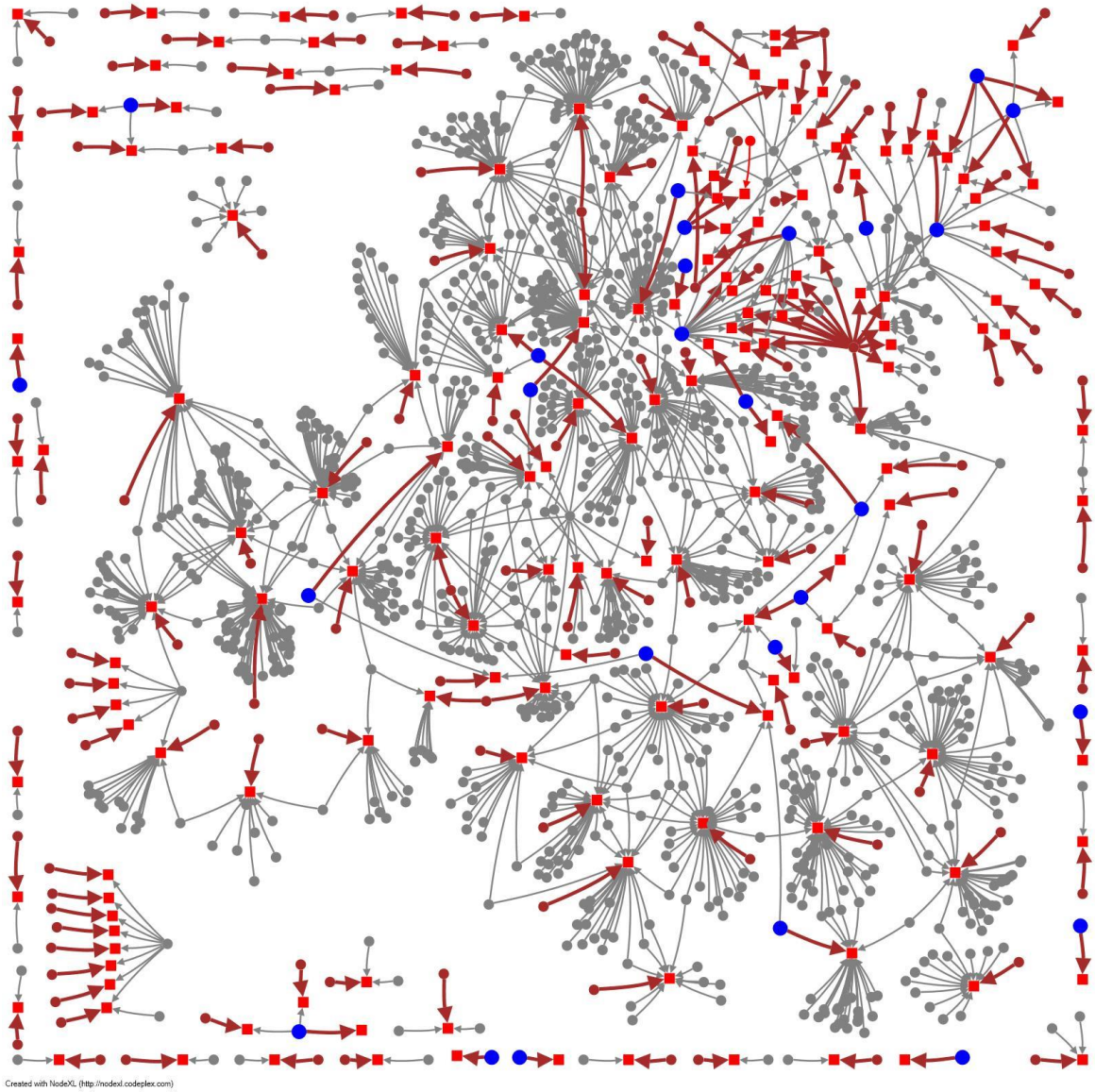

Fig. 8. Bimodal network of liked posts and users, C# group. Vertices: red squares – posts; circles – users, including: grey – only likers, brown – only posters, blue – likers-&-posters. Edges: grey – liking, brown – post authorship. Non-liked posts and inactive users not included. Geolabels omitted due to the network size.

Co-like network demonstrates a similar tendency. Since many likers are deactivated, edges of which the cities of both connected vertices were known numbered to only 569 out of 1 595 (36%); they embraced 554 vertices. Among those, there were eight self-likes, and only 22 were same-city likes, 19 of which were within Moscow. This network, unlike commenting network, was dominated by popularity stars (like-receivers), while super-active like-givers were much fewer and much less manifest (see Fig. 7). Most often, all or a lion share of likes were given to the same post of a user, which means that this was a network of occasional visitors leaving their only like (or their 2-3 likes) for an attractive message, not for its author. This latter trend can better be seen in the bimodal network which shows all posts ever liked (Fig. 8): only one user can be observed that got likes to really a multitude of posts: those were the posts written on behalf of the group itself (i.e. by its moderator).

Conclusion

What conclusions can be drawn from the case study of an SNS software developers' group and a few of its immediate neighbors that share the top positions in the search results? Such conclusions, of course, are limited, still the data give us a number of insights. First, all leading developers' VK groups are joined by people regardless of their location: the distribution of members by city is close to that of VK in general. Furthermore, members develop much more friendships with people from other cities than with people of their location. In our small sample, the number of people per city decreases with the decrease of the group size, and although we do not know whether this is a coincidence, the data suggests that this tendency will most probably persist in the long tail, but will be supplemented so that it will resemble the distribution of within-city densities (Fig. 4a). That is, the proportion of the same-city friendships will drop in one portion of groups, but another type of locally oriented groups will emerge. Actually, it does show up in the search results at the level of 2 000 members and below where one can find groups that indicate location in their very title (Kiev, St.Petersburg, Krasnoyarsk, Crimea being among the largest groups). It is thus may be concluded that there do exist SNS groups that fulfil a function of connecting across space.

Why would people in a professional community of software developers like to connect across space? A general purpose SNS, like VK, is not suitable for collaborative software development or for following the newest trends, since many "lay" persons actually come and post there. As we have seen in our case study, no ongoing projects or occupational trends are being discussed, so *awareness* and *collaboration* functions are absent. However, *learning* in the form of advice-seeking

/ giving, and *networking* for finding jobs or freelance work and getting software development services are present in abundance. More specific dialogue and problem-solving may be expected in those communities that rely on thread discussions more than on the wall posting. Anyway, the observed communication also develops across city boundaries.

References

1. Backstrom, L., Boldi, P., Rosa, M., Ugander, J., Vigna, S. (2012) Four degrees of separation. *Proceedings of the 3rd Annual ACM Web Science Conference*, New York, pp. 33-42.

2. Barcellini F., Détienne F., Burkhardt J.-M., Sack W. (2005) A Study of Online Discussions in an Open-Source Software Community. *Communities and Technologie*s 2005, pp. 301-320.

3. Bryant E.M., Marmo, J. (2012).The rules of Facebook friendship: A two-stage examination of interaction rules in close, *Journal of Social and Personal Relationships.* 0265407512443616.
4. Cairncross, F., 1997. *The Death of Distance: How the Communications Revolution is Changing Our Lives*. Harvard Business School Press, Boston.

5. Caren, N., Gaby, S. (2011) Occupy Online: Facebook and the Spread of Occupy Wall Street. *SSRN Electronic Journal*. Doi: 10.2139/ssrn.1943168.

6. Catanese, S., De Meo, P, Ferrara, E., Fiumara, G. (2010) Analyzing the Facebook Friendship Graph. Available at: http://www.emilio.ferrara.name/wp-content/uploads/2011/06/fb-rel-analysis-MIFI2010-fin.pdf (accessed April 1, 2014).
7. Duncan-Howell, J. (2010), Teachers making connections: Online communities as a source of professional learning. *British Journal of Educational Technology*, 41: 324–340. doi: 10.1111/j.1467-8535.2009.00953.x

8. Goldenberg J., Moshe L. (2009) Distance Is Not Dead: Social Interaction and Geographical Distance in the Internet Era, arXiv:0906.3202 [cs.CY], last revised 19 Oct 2009.

9. Goos, M.E., Bennison A. (2008). Developing a communal identity as beginning teachers of mathematics: Emergence of an online community of practice. *Journal of Mathematics Teacher Education*, February 2008, Vol. 11, Issue 1, pp. 41-60.
10. Hara, N., Hew, K.F. (2007) Knowledge-sharing in an online community of health-care professionals, *Information Technology & People*, Vol. 20 Iss: 3, pp.235 – 261

11. Huang, V.C.X. (2012). Professional commitment, organizational commitment, and the intention to leave for professional advancement, *Information Technology & People*, Vol. 25 Iss 1 pp. 31 – 54.

12. Lasker J.N., Sogolow E.D. Sharim R.R. (2005) The Role of an Online Community for People With a Rare Disease: Content Analysis of Messages Posted on a Primary Biliary Cirrhosis Mailinglist. *Journal of Medical Internet Research*, 2005 Jan-Mar 7 (1): e10.

13. Leskovec J., Lang K.J., Dasgupta, A., Mahoney M.W. (2008), Statistical properties of community structure in large social and information networks, *WWW '08 Proceedings of the 17th international conference on World Wide Web*, Beijing, China: ACM, pp. 695-704.
14. Marks, A., Scholarios, D. (2007) Revisiting technical workers: professional and organizational identities in the software industry. *New Technology, Work and Employment* 22:2, pp. 98-117.
15. Porter C.E. (2004), A Typology of Virtual Communities: A Multi-Disciplinary Foundation for Future Research, *Journal of Computer-Mediated Communication*, Volume 10, Issue 1, p. 00.
16. Rheingold H. (2000), *The Virtual Community: Homesteading on the Electronic Frontier*. Revised Edition. Cambridge, Mass.: MIT University Press.
17. Scellato, S., Mascolo, C., Musolesi, M., Latora, V. (2010) Distance matters: geo-social metrics for online social networks. *Proceedings of WOSN'10 Proceedings of the 3rd conference on Online social networks*. Available at: http://static.usenix.org/event/wosn/tech/full_papers/Scellato.pdf (accessed April 1, 2014)

18. Singer, L., Filho F.F., Storey M.-A. (2014) Software engineering at the speed of light: how developers stay current using twitter. *ICSE 2014 Proceedings of the 36th International Conference on Software Engineering*, New York, pp. 211-221.

19. Smith M.A., Kollock P. (1999), *Communities in Cyberspace*. Routledge.
20. Takhteyev, Y.; Gruzd, A. & Wellman, B. (2011), Geography of Twitter networks, *Social Networks* 34, pp. 73-81.
21. Thompson L.A., Dawson K., Ferdig R., Black E.W., Boyer J., Coutts J., Black N.P. (2008) The Intersection of Online Social Networking with Medical Professionalism. *Journal of General Internal Medicine*. July 2008, Volume 23, Issue 7, pp. 954-957.
22. Traud, A. L.; Mucha, P. J.; Porter, M. A. & Porter, M. A. (2011), Social Structure of Facebook Networks, *Physica A* 391, pp. 4165–4180.
23. Wickramasinghe, V., Weliwitigoda, P. (2011) Benefits gained from dimensions of social capital: a study of software developers in Sri Lanka, *Information Technology & People*, Vol. 24 Iss: 4, pp. 393 – 413

24. Wu, Y., Kropczynski J., Shih P.C., Carroll J.M. (2014). Exploring the ecosystem of software developers on GitHub and other platforms. *CSCW Companion '14 Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing*, New York, pp. 265-268.

25. Демьяненко И. Статистика по профилям пользователей ВКонтакте. Available at: http://habrahabr.ru/post/123856/ (accessed on April 1, 2014).

Olessia Y. Koltsova

National Research University Higher School of Economics
Director of the Laboratory of Internet Research
E-mail: olessia.koltsova@gmail.com
Tel.: +7 (812) 677-94-52

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.