

# Indicators of Connectivity for Urban Scientific Communities in Russian Cities

Fedor Krasnov<sup>1</sup>, Rostislav E. Yavorskiy<sup>2</sup>, and Evgeniya Vlasova<sup>2</sup>(✉)

<sup>1</sup> Skolkovo Foundation, Moscow, Russia  
fk@sk.ru

<sup>2</sup> Higher School of Economics, National Research University, Moscow, Russia  
ryavorsky@hse.ru, vlasova.eug@gmail.com

**Abstract.** In this paper we study formal indicators of connectivity for a city community of researchers in Informatics and Cybernetics. The analysis is based on data available at the Scientific Electronic Library portal <http://eLibrary.ru>. Starting from the co-authorship relation we construct connectivity graph for research institutions for all major Russian cities and suggest using size of the 2-core component to measure the connectivity of the local communities.

**Keywords:** Social network analysis · Research communities · Communities connectivity · Co-authorship relations

## 1 Introduction

### 1.1 Role of Collaboration in Creative Cities and Islands of Innovation

As it is stated in [1] “the *creative city* became the new hot topic among urban policymakers, planners, and economists, especially in North America and Western Europe”. Needless to say that initiative of Russian government to build Skolkovo Innovation Center also follows that trend. It is stressed out in [1] that among other factors creative cities have always been associated with free exchange of scientific ideas, which naturally raises the task of developing measurable indicators of that parameter. Monograph on intelligent cities [2] mentions science and technology parks as one of five category of island of innovation and lists *collaboration* between universities and businesses as the first factor of the productive environment. So it becomes quite important to develop methods and tools for measuring the collaboration level as according to Lord Kelvin “If you cannot measure it, you cannot improve it.”

### 1.2 How to Measure a City

In [3] fifteen indicators divided in five categories were selected to design index system of innovative city. Although the index system list contains several measures

of the local R&D community such as “Personal quantity engaged in R&D per million labor forces”, “The proportion of R&D fund to GDP”, “R&D personal full-time equivalent” and others, none of the indicators evaluates the collaboration level in the considered city.

The task of local communities ranking is quite similar to the well known problem of university ranking, see [10] and [11] for a detailed study on the subject. Another project which is closely related to our research is Map of Russian Science, which is currently in trial operation phase, see [13] and [14] for more.

Our general approach to the study of urban professional communities was described in [4]. It was based on mathematical model of a community as a dynamic socio-semantic network described in [6], see also [9]. A more detailed overview of works on scientific collaboration could be found in [7] and [8]. Some other factors of professional online communities are studied in [5].

### 1.3 Maturity Measures of Professional Community

In [4] an approach was proposed for assessing an urban professional IT community maturity level. It is based on measuring two sets of parameters that characterize the level of competence and the density of the network of contacts. With such approach, the formal model is described in terms of combination of social and semantic networks. The study also provides results of the pilot testing of the proposed approach for assessment of several city-wide IT-communities in central Russia.

The proposed rating of a professional community provides a system approach to assessing the current status of professional communities. However, this approach (in its current form) has a certain limited scope of applicability. First of all, a more definitive list of parameters and their weight needs to be elaborated, specifically to account for financial performance and social/demographic data of a region, information on registered legal entities, etc. In addition, a system of scoring by experts should be replaced with the one automatical, based on social/demographic and other data to minimize the human subjectivity factor.

According to [4] the key factors defining professional IT communities maturity level are competences and contacts. The both factors can be decomposed into four components. To these two groups of factors we add the third component, activity bonus, including positive factors that don't fit in the previous ones. Total 100 points:

1. Competencies (max – 40 points)
  - (a) Development of IT education – 10 points
  - (b) Development of IT industry – 10 points
  - (c) Development of Business Education – 10 points
  - (d) Research in Computer Science – 10 points
2. Contacts (max – 40 points)
  - (a) Regular IT conferences and workshops – 10 points
  - (b) Web communities, blogs and forums targeted at IT audience – 10 points
  - (c) Groups in Social networks – 10 points
  - (d) Focused IT Media – 10 points

### 3. Activity bonus – (max 20 points)

In this paper we assume that the academic community maturity level index can be either decomposed into the level of competence and the density of the network of contacts.

#### 1.4 Goal of This Paper

As it was mentioned before, in the original study the scoring was performed by the group of independent experts. In this paper we suggest an automated procedure based on publicly available data on publications to measure the connectivity level of Russian major cities research computer science communities. We take into account both absolute number of scientific publications and links with the other scientific centers, which are determined as the number of co-authored papers.

## 2 The Dataset

Our analysis is based on data available at <http://elibrary.ru>, which is the largest Russian scientific portal, aggregating works on science, technology, medicine, and education. It contains over 18 million of articles and publications from more than 3200 Russian journals, see [12].

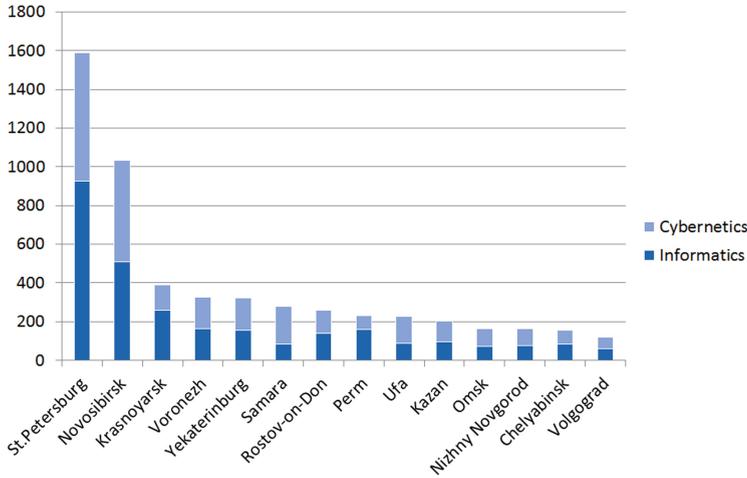
### 2.1 The Search Criteria

We restricted the focus of our study to Russian cities with population over 1 million citizens [15] and less than 5 to exclude cities-multimillionaires widely differing from the cities we consider in this paper by its size and structure.

The search configuration was the following.

1. Start with the extended search form.
2. Put the city name into the search field. The Russian cities-millionaires we studied are Novosibirsk, Yekaterinburg, N. Novgorod, Kazan, Samara, Omsk, Chelyabinsk, Rostov-on-Don, Ufa, Volgograd, Krasnoyarsk, Perm, and Voronezh.
3. Indicate that the search should run through the authors affiliations, tick on all types of publications.
4. Restrict the scientific area to Cybernetics (28.00.00) or Informatics (20.00.00).
5. Switch on the morphology trigger.
6. Specify the publication years, 2011 till 2013.

Having followed these instructions we downloaded 26 collections (13 cities, 2 scientific areas) with more than half a hundred articles in each. The exact number of articles on Cybernetics and Informatics published in the last 3 years according to the Electronic Library portal is provided in Fig. 1.



**Fig. 1.** Rating of major Russian cities (without Moscow) with respect to the number of 2011–2013 publications in Informatics and Cybernetics according to data from the Electronic Library portal <http://eLibrary.ru>

### 3 Research Organizations Connectivity Graph

Each paper from the data set contains information about the authors and their affiliations. At first, we extracted the list of the institutions from the considered city, whose members have published at least one research paper since 2011. As it was mentioned before, in this paper we restrict our attention to IT related areas, namely, Informatics and Cybernetics.

These institutions are the nodes of the *research organizations connectivity graph* for a city. Two institutions are connected if there is at least one paper in the data set which is co-authored by people from these institution.

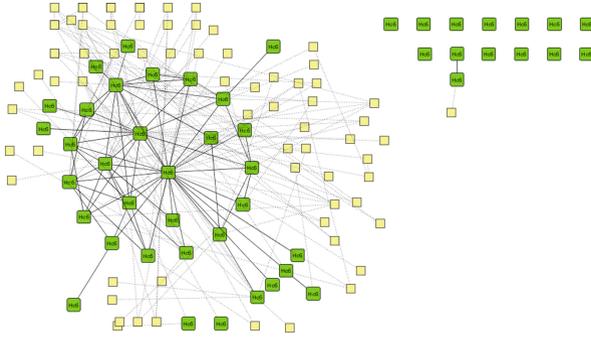
Some of the papers may be co-authored by researchers from the different cities. So, to make the picture complete we also show connections to the other cities institutions on the graph. See the graph for Novosibirsk in Fig. 2.

To make the picture less noisy we remove isolated nodes and pairs. Also we stress out all the edges of the core part with bold lines. The resulting graph for Novosibirsk is given in Fig. 3.

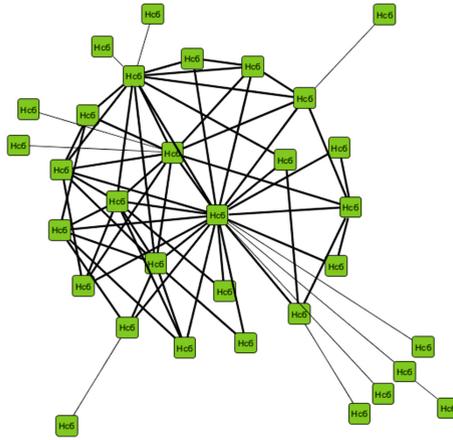
The formal definitions follow.

We say that two different organizations  $v_1$  and  $v_2$  *collaborate* if there is at least one paper in the dataset co-authored by employees from the both  $v_1$ ,  $v_2$  (and maybe some other organizations). Then, *the research organizations connectivity graph*  $G = (V, R)$  for a given city has the following parameters.

- $V$  is a set of nodes. Each node denotes an organization.  $V$  consists of the following two disjoint parts:



**Fig. 2.** Research organizations connectivity in Novosibirsk



**Fig. 3.** Refined graph for Novosibirsk

- $V_{in}$  comprises all organizations from the considered city such that there exists at least one paper in the dataset, which is authored or co-authored by someone working at this organization;
  - $V_{out}$  denotes all organizations outside the city, which collaborate with some organizations in  $V_{in}$ .
- $R$  is irreflexive symmetric binary relation on  $V$  which links different collaborating organizations. So,  $\forall v \neg R(v, v)$  and  $\forall u, v (R(u, v) \rightarrow R(v, u))$ .

## 4 Analysis

For each considered city all the nodes in  $V_{in}$  are naturally classified into the following six groups:

- $L_0$  is the subset of isolated nodes.
- $L_1$  denotes isolated nodes with external links.

- $L_2$  stands for isolated pairs, i. e. pairs of collaborating institutions from the considered city, that may have some connectors from the other cities, but cannot have more connectors from the considered one.
- $L_3$  stands for dangling nodes belonging to a larger connected component.
- $L_4$  includes nodes on dangling paths.
- $L_5$  nodes from the graph 2-core.

The detailed definitions of the layers are given below.

**Isolated Nodes,  $L_0$ .** It turned out that each city has quite a big number of organizations with no connections at all (Fig. 4).

$$L_0 = \{v \in V_{\text{in}} \mid \forall w \neg R(v, w)\} \quad (1)$$

**Isolated Nodes with External Links,  $L_1$ .** This group consists of institutions, which are not connected with other organizations in the city, but have collaborators in some other city (Fig. 5).

$$L_1 = \{v \in V_{\text{in}} \mid \exists w \in V_{\text{out}} R(v, w) \wedge \forall u \in V_{\text{in}} \neg R(v, u)\} \quad (2)$$

**Nodes in Collaborating Pairs,  $L_2$ .** A collaborating pair consists of two connected institutions with no links to other organizations in the city.

$$L_2 = \bigcup \{v_1, v_2 \in V_{\text{in}} \mid R(v_1, v_2) \wedge \forall u \in V_{\text{in}} \setminus \{v_1, v_2\} (\neg R(v_1, u) \wedge \neg R(v_2, u))\} \quad (3)$$

**Dangling Nodes at Bigger Connected Components,  $L_3$ .** Organizations linked to a larger connected component that consists of more than 2 nodes.

$$L_3 = \{v \in V_{\text{in}} \mid \exists! w \in V_{\text{in}} (R(v, w) \wedge \exists u (u \neq v \wedge R(u, w)))\} \quad (4)$$

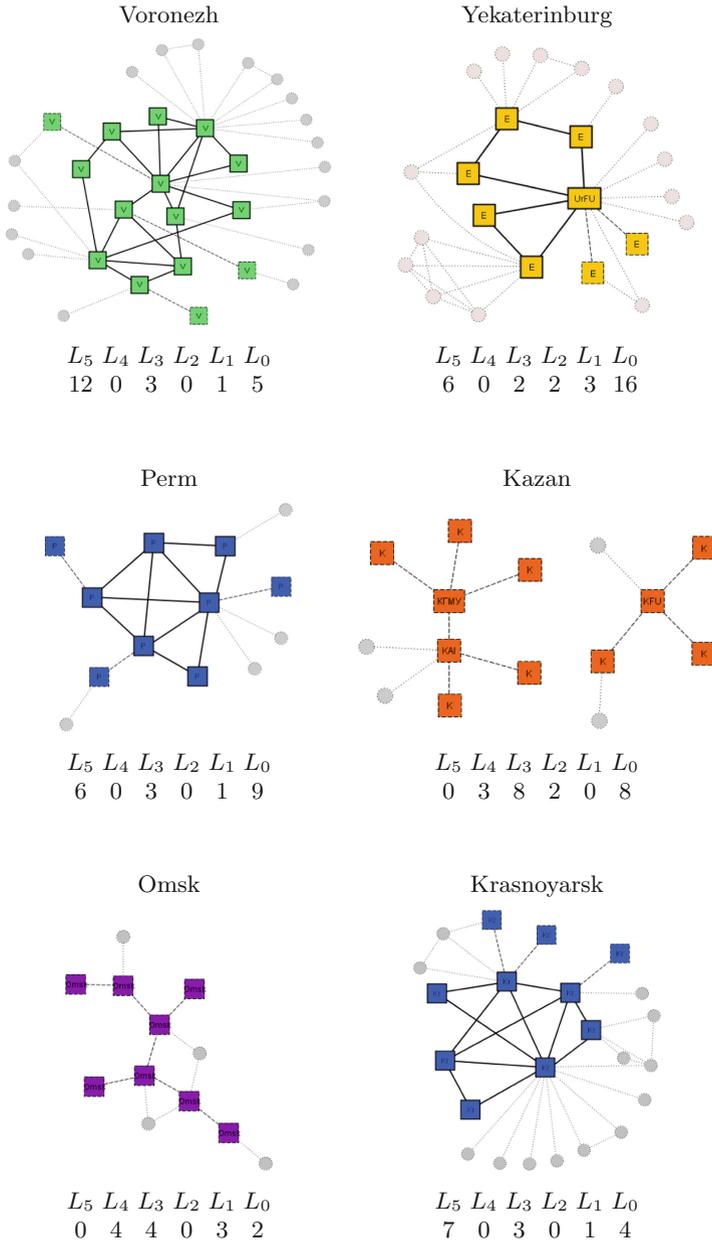
**Nodes on Dangling Paths,  $L_4$  and the Graph 2-Core,  $L_5$ .**  $L_4$  stands for organizations linked to a connected component via one single path. This class includes nodes which have more than 1 neighbor (so they are not in  $L_3$ ), but does not belong to graph 2-core, which is  $L_5$ .

Finally,  $L_5$  is the graph 2-core, the maximal subgraph with minimum degree at least 2. This group includes all groups of several collaborating institutions with two or more collaborators each. (Sometimes 2-core is defined as a maximal connected subgraph where every node has at least two neighbours, the connectivity is not required here.) In general the connectivity analysis would require computing  $n$ -core for  $n > 2$ , yet for the given dataset these are empty for most of the cities.

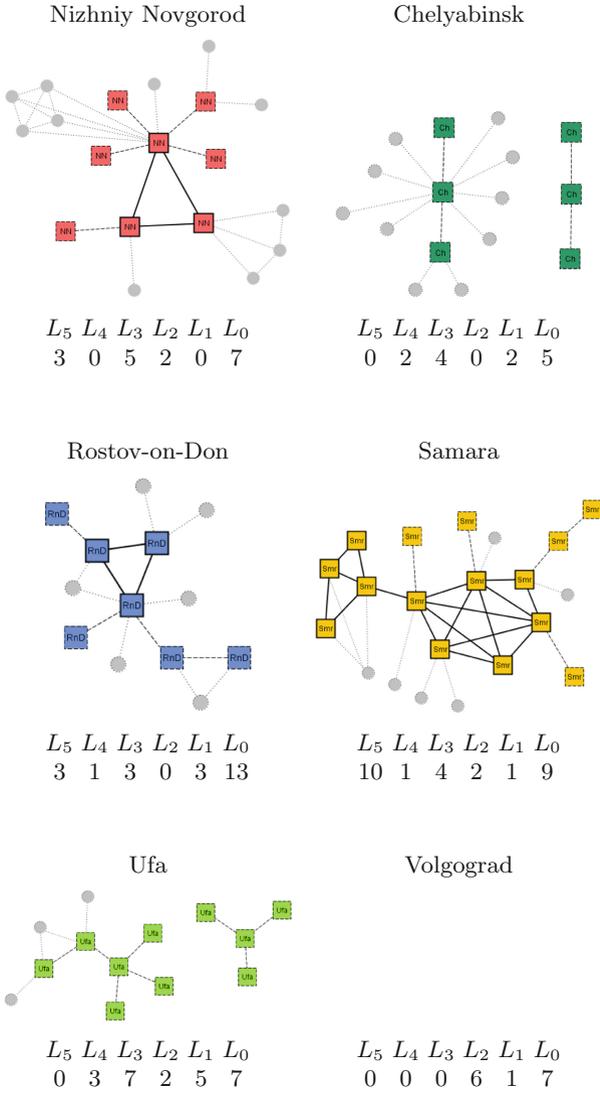
$$L_4 = \max\{U \subseteq V_{\text{in}} \mid \forall u \in U \exists v, w \in U (v \neq w \wedge R(u, v) \wedge R(u, w))\} \quad (5)$$

And then

$$L_5 = \{v \in (V_{\text{in}} \setminus L_4) \mid \exists u, w (u \neq w \wedge R(v, u) \wedge R(v, w))\} \quad (6)$$



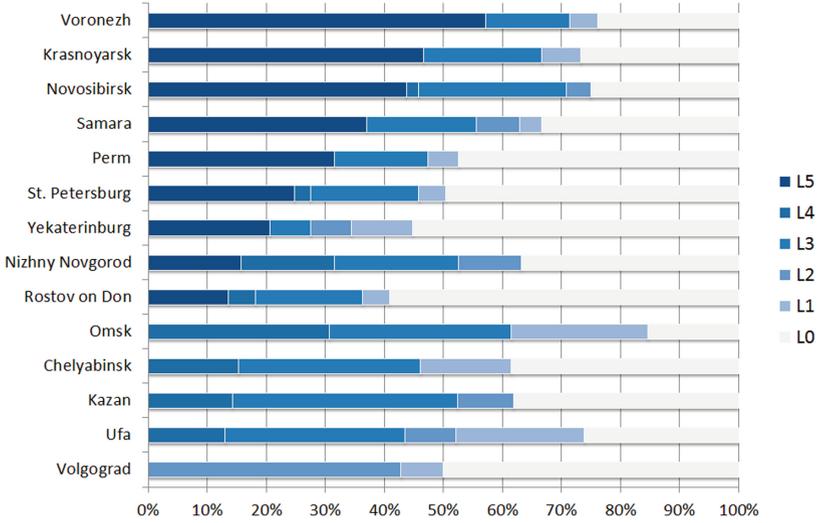
**Fig. 4.** Core part of the connectivity graphs for major Russian cities. Part 1. (Nodes from  $L_0$ ,  $L_1$ ,  $L_2$  are not displayed.)



**Fig. 5.** Core part of the connectivity graphs for major Russian cities. Part 2. (Nodes from  $L_0$ ,  $L_1$ ,  $L_2$  are not displayed.)

## 5 The Final Rating of Russian Cities-Millionaires Research Communities with Respect to Their Levels of Connectivity

The level of connectivity is one of the key points of measuring the community maturity level.



**Fig. 6.** The final rating of cities with respect to the connectivity of the research community based on the co-authorship relation for papers in Cybernetics and Informatics during 2011–2013

In this paper we suggest a formal procedure that helps computing the connectivity graph  $G$  for any research community, split its vertexes into several connectivity levels  $L_0, \dots, L_5$  defined above. So, each city gets the following vector of the normalized connectivity characteristics:

$$l = (l_5, l_4, l_3, l_2, l_1, l_0), \text{ where } l_i = \frac{|L_i|}{|V_{in}|} \text{ and } \sum_{i=0}^5 l_i = 1. \quad (7)$$

According to the connectivity graphs we build the final rating of Russian cities based on lexicographical order of the vectors of normalized connectivity characteristics. See Fig. 6.

## 6 Conclusion

The idea of measuring urban communities was discussed in many papers and has shown its ambiguity, see e.g. [16]. Understanding the structure of urban communities considered in this paper brings us closer to finding the key phases in the development of urban communities and to the understanding of phase transitions.

Although our study at this stage aims at identifying explicit phases of development of urban communities, we believe that the observed characteristics of urban communities allow us to make a step towards building a descriptive system and individual classification of the urban communities.

## References

1. Andersson, D.A., Andersson, A.E., Mellander, C.: Handbook of Creative Cities. Edward Elgar Publishing, Cheltenham (2011)
2. Komminos, N.: Intelligent Cities: Innovation, Knowledge Systems and Digital Spaces. Routledge, London (2013)
3. Guo, P., Chen, J., Song, Z.: Construction of evaluation system of innovative city and empirical research. *J. Appl. Sci.* **13**(22), 5428–5433 (2013)
4. Krasnov, F., Yavorskiy, R.: Measurement of maturity level of a professional community. *Bus. Inf.* **1**(23), 64–67 (2013) (in Russian). [http://www.bijournal.hse.ru/en/2013--1\(23\)/86058013.html](http://www.bijournal.hse.ru/en/2013--1(23)/86058013.html)
5. Krasnov, F., Ustalov, D., Yavorskiy, R.: Comparison of online communities on the base of lexical analysis of the news feed. In: Proceedings of 2-nd Conference on Analysis of Images, Networks and Texts, Yekaterinburg, pp. 254–257, 4–6 April 2013 (in Russian)
6. Yavorsky, R.: Research challenges of dynamic socio-semantic networks. In: CEUR Workshop Proceedings, vol. 757 (2011)
7. Taramasco, C., Cointet, J.-P., Roth, C.: Academic team formation as evolving hypergraphs. *Scientometrics* **85**(3), 721–740 (2010)
8. Roth, C., Wu, J., Lozano, S.: Assessing impact and quality from local dynamics of citation networks. *J. Inf.* **6**(1), 111–120 (2012)
9. Roth, C.: Socio-semantic frameworks. *Adv. Complex Syst.* **16**(04–05), 1350013 (2013)
10. van Vught, F.A., Ziegele, F.: U-Multirank: design and testing the feasibility of a multidimensional global university ranking. Final report (2011). <http://doc.utwente.nl/85192/1/Vught11umulti.pdf>
11. van Vught, F.A., Ziegele, F.: Multidimensional Ranking: The Design and Development of U-Multirank, vol. 37. Springer, Heidelberg (2012)
12. Scientific Electronic Library. <http://eLibrary.ru>
13. Polyakov, A.M.: Map of Russian science. First results. Conference SCIENCE INDEX 2013: Analytical Instruments and Services for Assessment of Scientific Work. MGIMO, Moscow, 9–10 December 2013 (in Russian). [http://www.elibrary.ru/projects/science\\_index/conf/2013/presentations/polyakov.pdf](http://www.elibrary.ru/projects/science_index/conf/2013/presentations/polyakov.pdf)
14. Map of Russian Science. <http://www.mapofscience.ru>
15. Russian Federation Federal State Statistics Service. [http://www.gks.ru/bgd/regl/b10\\_109/Main.htm](http://www.gks.ru/bgd/regl/b10_109/Main.htm)
16. Cox, A.: What are communities of practice? A comparative review of four seminal works. *J. Inf. Sci.* **31**(6), 527–540 (2005)