Федеральное государственное автономное образовательное учреждение высшего образования
"Национальный исследовательский университет
"Высшая школа экономики"

Факультет компьютерных наук
Департамент анализа данных

Рабочая программа дисциплины «Анализ сетевых структур» (на английском языке) «Network Science»

для образовательной программы «Науки о данных»
направления подготовки 01.04.02. Прикладная математика и информатика
уровень - магистр

Разработчик программы
Л.Е. Жуков, PhD, профессор, lzhukov@hse.ru
И.А. Макаров, iamakarov@hse.ru

Одобрена на заседании департамента анализа данных и искусственного интеллекта
«___»____________ 2018 г.
Руководитель департамента анализа данных и искусственного интеллекта Школы
С.О. Кузнецов ________

Утверждена Академическим советом образовательной программы
«___»____________ 2018 г., № протокола_________________

Академический руководитель образовательной программы
С.О. Кузнецов _________________

Москва, 2018

1. Teachers

Author, lecturer: Leonid Zhukov, National Research University Higher School of Economics, Department of Data Analysis and Artificial Intelligence, professor

Tutor: Ilya Makarov, National Research University Higher School of Economics, Department of Data Analysis and Artificial Intelligence, senior lecturer, deputy head

2. Scope of Use

The present program establishes minimum demands of students' knowledge and skills, and determines content of the course.

The present syllabus is aimed at department teaching the course, their teaching assistants, and students of the Master of Science program 010402 «Data Sciences»,

This syllabus meets the standards required by:

- Educational standards of National Research University Higher School of Economics;
- Educational program «Data Science» of Federal Master's Degree Program 010402 «Applied Mathematics and Informatics», 2015;
- University curriculum of the Master's program in «Data Science» for 2015.

3. Summary

The course "Network Science" introduces students to new and actively evolving interdisciplinary field of network science. Started as a study of social networks by sociologists, it attracted attention of physicists, computer scientists, economists, computational biologists, linguists and others and become a truly interdisciplinary field of study. In spite of the variety of processes that form networks, and objects and relationships that serves as nodes and edges in these networks, all networks poses common statistical and structural properties. The interplay between order and disorder creates complex network structures that are the focus of the study. In the course we will consider methods of statistical and structural analysis of the networks, models of network formation and evolution and processes developing on network. Special attention will be given to the hands-on practical analysis and visualization of the real world networks using available software tools and modern programming languages and libraries.

4. Learning Objectives

Learning objectives of the course "Network Science" are to familiarize students with a new rapidly evolving filed of network science, and provide practical knowledge experience in analysis of real world network data.

5. Learning outcomes

After completing the study of the discipline "Network Science", the student should:

- Know basic notions and terminology used in network science
- Understand fundamental principles of network structure and evolution
- Learn to develop mathematical models of network processes
- Be capable of analyzing real world network data

After completing the study of the discipline "Network Science" the student should have the following competences:

Competence	Code	Code (UC)	Descriptors (indicators of achievement of the result)	Educative forms and methods aimed at generation and development of the competence
The ability to reflect developed methods of activity.	SC-1	SC-M1	The student is able to reflect developed mathematical methods to network sciences	Lectures and tutorials, group discussions, presentations, paper reviews.
The ability to propose a model to invent and test methods and tools of professional activity	SC-2	SC-M2	The student is able to improve and develop research methods of linear optimization, approximation and computational problem solvation.	Classes, home works.
Capability of development of new research methods, change of scientific and industrial profile of self-activities	SC-3	SC-M3	The student obtain necessary knowledge in network science, which is sufficient to develop new methods on other sciences	Home tasks, paper reviews,
The ability to describe problems and situations of professional activity in terms of humanitarian, economic and social sciences to solve problems which occur across sciences, in allied professional fields.	PC-5	IC-M5.3_5.4_5.6_2.4.1	The student is able to describe network problems problems in terms of computational mathematics.	Lectures and tutorials, group discussions, presentations, paper reviews.
The ability to detect, transmit common goals in the professional and social activities	PC-8	SPC-M3	The student is able to identify mathematical aspect in network research evaluate correctness of the used methods, and their applicability in each current situation	Discussion of paper reviews; cross discipline lectures

6. Place of the discipline in the Master's program structure

The course "Network Science" is a course taught in the first year of the Master's program 010402 "Data Sciences" and is a base course for specialization "Intelligent Systems and Structural Analysis"

Prerequisites

The course is based on knowledge and understanding of

- Discrete mathematics
- Algorithms and data structures
- Linear algebra
- Theory of probability and statistical analysis

It also requires some programming experience in one of programming languages:

- Python
- Matlab
- R

7. Schedule

One pair consists of 1 academic hour for lecture and 1 academic hour for classes after lecture.

№	Topic	Total hours	Contact hours		Self-study
			Lectures	Seminars	
1	Introduction to network science	9	3	2	4
2	Power laws	13	3	2	8
3	Models of network formation	15	3	4	8
4	Structure, nodes and links analysis	17	3	6	8
5	Network communities	17	3	6	8
6	Diffusion and epidemics on networks	15	3	4	8
7	Influence propagation	17	3	6	8
8	Information cascades	15	3	4	8
9	Evolving networks and link prediction	17	3	6	8
10	Network visualization	17	3	6	8
	Total	152	30	46	76

8. Requirements and Grading

	Type of work	Characteristics		
		3	4	
Type of grading	Homework	4	4	Solving homework tasks and examples.
	Special homework – research projects	1	1	Independent modelling and verification of research papers results
	Mid Term Exam	1		
	Final Exam		1	Written exam

9. Assessment

The assessment consists of weekly classwork and homework. Short quizzes are assigned after each lecture through the google-forms to make students pass their homeworks (otherwise the maximal grade for HW will not exceed 4 of 10). Students have to demonstrate their knowledge in each lecture topic concerning both theoretical facts, and practical tasks' solving. All tasks are connected through the discipline and have increasing complexity. The two practical projects should be presented in class and graded by both, lecturers and students. The weight of a project is three times greater than the weight of each HW.

Each technical report should be made in IPython Notebook format.

Labs are to be uploaded in DataJoy Group storage.

Students obtain the following points per each condition:

- Correct data analysis – 4 points;
- Suitable visualization – 3 points;
- Report completeness and formulas' descriptions – 2 points;
- Extra 1 point for overall report evaluation and excellent research and soft skills.

Technical details for reports and projects are placed to the corresponding technical section.

Final assessment is the final exam. Students have to demonstrate knowledge of theory facts, but the most of tasks would evaluate their ability to solve practical examples, present straight operation, and recognition skills to solve them.

The grade formula:

The exam will consist of 10 problems, giving 10 points each, total 100 points for the exam

Final course mark *is obtained from the following formula*:

$$O_{final} = 0,6*O_{cumulative}+0,4*O_{exam}.$$

The grades are rounded in favour of examiner/lecturer with respect to regularity of class and home works. All grades, having a fractional part greater than 0.5, are rounded up.

Table of Grade Accordance

Ten-point	Five-point	

Grading Scale	Grading Scale	
1 - very bad 2 – bad 3 – no pass	Unsatisfactory - 2	**FAIL**
4 – pass 5 – highly pass	Satisfactory – 3	**PASS**
6 – good 7 – very good	Good – 4	
8 – almost excellent 9 – excellent 10 – perfect	Excellent – 5	

10. Course Description

The following list describes main topics covered by the course with lecture order.

Topic 1. Introduction to network science

Content:
Introduction to new science of networks. Complex networks theory. Basic network properties and metrics. Networks examples.

Recommended reading:
1. Chapters 1-3 of Mark Newman. "Networks: An Introduction". Oxford University Press, 2010.
2. Chapters 1,2 of David Easley and John Kleinberg. "Networks, Crowds, and Markets: Reasoning About a Highly Connected World." Cambridge University Press 2010.

Supplementary reading:
1. Albert-Laszlo Barabasi and Eric Bonabeau. Scale Free Networks. Scientific American, p 50-59, 2003
2. Mark Newman. The physics of networks. Physics Today,2008
3. Stanley Milgram. The Small-World Problem. Psychology Today, Vol 1, No 1, pp 61-67, 1967
4. J. Travers and S. Milgram. An Experimental Study of the Small World Problem. Sociometry, vol 32, No 4, pp 425-433, 1969
5. Mark Granovetter. The strengtth of weak ties , American Journal of Sociology, 78(6):1360-1380, 1973.

Topic 2. Power laws

Content:

Power law distribution. Scale-free networks. Pareto distribution, normalization, moments. Zipf's law. Rank-frequency plot.

Recommended reading:
1. Chapter 8 of Mark Newman. "Networks: An Introduction". Oxford University Press,

2010.
2. Chapter 18 of David Easley and John Kleinberg. "Networks, Crowds, and Markets: Reasoning About a Highly Connected World." Cambridge University Press 2010.

Supplementary reading:
1. M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. Contemporary Physics 46(5), 323-351, 2005
2. Clauset, C.R. Shalizi, M.E.J. Newman. Power-law distributions in empirical data. SIAM Review 51(4), 661-703, 2009
3. M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. Internet Mathematics, vol 1, No. 2, pp. 226-251, 2004
4. M.L. Goldstein, S.A. Morris, and G.G. Yen. Problems with fitting to the power-law distribution, Eur. Phys. J. B 41, pp 255–258, 2004.

Topic 3. Models of network formation

Content:

Erdos-Reni random graph model. Poisson and Bernulli distributions. Distribution of node degrees. Phase transition, gigantic connected component. Diameter and cluster coefficient. Configuration model. Barabasi-Albert model. Preferential attachment. Equation in continues approximation. Time evolution of node degrees. Node degree distribution. Average path length and clustering coefficient. Small world model. Watts-Strogats model. Transition from regular to random. changes in clustering coefficient and average path length.

Recommended reading:
1. Chapters 12, 14 of Mark Newman. "Networks: An Introduction". Oxford University Press, 2010.
2. Chapter 20 of David Easley and John Kleinberg. "Networks, Crowds, and Markets: Reasoning About a Highly Connected World." Cambridge University Press 2010.

Supplementary reading:
1. P. Erdos and A. Renyi. On random graphs I. Publ. Math. Debrecen, 1959.
2. P. Erdos and A. Renyi. On the evolution of random graphs. Magyar Tud. Akad. Mat. Kutato Int. Koezl., 1960.
3. Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. . Nature 393:440-42, 1998.
4. AL Barabasi and R. Albert. Emergence of Scaling in Random Networks. Science, 286, 1999

Topic 4. Structure, nodes and links analysis

Content:

Node centrality metrics, degree centrality, closeness centrality, betweenness centrality, eigenvector centrality. Graph structure based metrics. PageRank, stochastic metric and Perron-Frobenius theorem. Power iterations. Hubs and Authorities. HITS algorithm. Kendall-Tau ranking distance. Structural equivalency metrics. Euclidean and Hamming distance. Correlation coefficient. Cosine similarity. Assortative mixing and homophily. Modularity. Mixing by degree.

Recommended reading:

1. Chapter 7 of Mark Newman. "Networks: An Introduction". Oxford University Press, 2010.
2. Chapter 14 of David Easley and John Kleinberg. "Networks, Crowds, and Markets: Reasoning About a Highly Connected World." Cambridge University Press 2010.

Supplementary reading:

1. Linton C. Freeman. Centrality in Social Networks. Conceptual Clarification. Social Networks, Vol 1, pp 215-239, 1978
2. Phillip Bonacich. Power and Centrality: A Family of Measures. American journal of sociology, Vol.92, pp 1170-1182, 1987.
3. S. Brin, L. Page. The PageRank Citation Ranknig: Bringing Order to the Web.
4. John M. Kleinberg. Authoritative Sources in a Hyperlinked Environment. Proc. 9th ACM-SIAM Symposium on Discrete Algorithms, 1998.
5. M. Newman. Mixing patterns in networks. Phys. Rev. E, Vol. 67, p 026126, 200

Topic 5. Network communities

Content:
Network communities. Graph density. Graph pertitioning. Min-cut, quotent and normalized cuts metrics. Divisive and agglomerative algorithms. Repeated bisection. Correlation matrix. Clustering. Edge Betweenness. Newman-Girvin algorithm. Spectral methods. Modularity maximization algorithm (Newman). Approximation algorithms. Randomized min-cut (Karges's algorithm). Probability of finding min-cut. Multilevel paradigm. Multilevel algorithm for power law graphs. Local clustering. Conductance. Nibble Algorithms. Graph motiffs, k-cores, diad and triad census

Recommended reading:

1. Chapter 11 of Mark Newman. "Networks: An Introduction". Oxford University Press, 2010.

Supplementary reading:

1. M.E.J. Newman, M. Girvan. Finding and evaluating community structure in networks. Phys. Rev. E 69, 026113, 2004.
2. M.E.J. Newman. Modularity and community structure in networks. PNAS Vol. 103, N 23, pp 8577-8582, 2006
3. S. E. Schaeffer. Graph clustering. Comp. Sci. Rev., Vol. 1, p 27-64, 2007
4. D.R. Karger. Global min-cuts in RNC, and other ramifications of a simple min-cut algorithm. Proceedings SODA '93, p. 21-30, 1993
5. A. Abou-rjeili, G. Karypis. Multilevel algorithms for partitioning power-law graphs. In Proceedings IPDPS '06, p 10, 2006
6. G.Karypis and V. Kumar. A fast and high quality multilevel scheme for partitioning irregular graphs. SIAM J. on Sci. Comp., Vol. 20, p 359-392, 1998.
7. Daniel A. Spielman, Shang-Hua Teng. A Local Clustering Algorithm for Massive Graphs and Its Application to Nearly Linear Time Graph Partitioning. SIAM Journal on computing, Vol. 42, p. 1-26, 2013
8. R. Andersen, F. Chung, K. Lang. Local graph partitioning using pagerank vectors. In Proc. FOCS, 2006.
9. S. Fortunato. Community detection in graphs . Physics Reports, Vol. 486, pp. 75-174, 2010

10. V. Batagelj, M. Zaversnik. An O(m) Algorithms for Cores Decomposition of Networks. 2003
11. L. da F. Costa, F. A. Rodrigues, et. al. Characterization of complex networks: A survey of measurements.Advances in Physics, Vol. 56, pp. 167-242, 2007
12. R. Milo, S. Shen-Orr, S. Itzkovitz et al. Network motifs: simple building blocks of complex networks. Science 298 (5594): 824–827, 2002

Topic 6. Diffusion and epidemics on networks

Content:

Physical diffusion. Diffusion equation. Diffusion on networks. Discrete Laplace operator, Laplace matrix. Solution of the diffusion equation. Random walks on graph. Epidemic models SI, SIS, SIR. Solution of diff. equations. Limiting cases. Modeling of infection propagation

Recommended reading:

1. Chapters 6, 17 of Mark Newman. "Networks: An Introduction". Oxford University Press, 2010.
2. Chapter 21 of David Easley and John Kleinberg. "Networks, Crowds, and Markets: Reasoning About a Highly Connected World." Cambridge University Press 2010.

Supplementary reading:

1. Lovasz, L. Random walks on graphs: a survey. In Combinatorics, Paul Erdos is eighty. pp. 353 – 397. Budapest: Janos Bolyai Math. Soc., 1993
2. Chung, Fan R.K. Spectral graph theory (2ed.). Providence, RI: American Math. Soc., 1997
3. H.W. Hethcote. The Mathematics of Infections Diseases. SIAM Review, Vol. 42, No. 4, pp. 599-653, 2000
4. Matt. J. Keeling and Ken.T.D. Eames. Networks and Epidemics models Journal R. Soc. Interface, Vol 2, pp 295-307, 2005
5. G. Witten and G. Poulter Simulations of infections diseases on networks Computers in Biology and Medicine, Vol 37, No. 2, pp 195-205, 2007
6. M. Kuperman and G. Abramson Small World Effect in an Epidemiological Model., Phys. Rev. Lett., Vol 86, No 13, pp 2909-2912, 2001

Topic 7. Influence propagation

Content:

Social diffusion. Granovetter's Threshold model of Collective behavior. Most influential nodes in network. Cascades in networks. Basic cascade model. Cascade capacity.

Recommended reading:

1. Chapter 19 of David Easley and John Kleinberg. "Networks, Crowds, and Markets: Reasoning About a Highly Connected World." Cambridge University Press 2010.

Supplementary reading:

1. Mark S. Granovetter. Threshold Models of Collective Behavior. American Journal of Sociology Vol. 83, No. 6, pp. 1420-1443, 1978.
2. H. Peyton Young. The Diffusion of Innovations in Social Networks. In L. E. Blume and S. N. Durlauf (eds.), The Economy as an Evolving Complex System III (2003)
3. D. Kempe, J. Kleinberg, E. Tardos. Maximizing the Spread of Influence through a Social Network. In Proc. KDD 2003.
4. D. Watts. A simple model of global cascades on random networks. Proc. Natl. Acad. Sci., vol. 99 no. 9, 5766-5771, 2002.
5. D. Kempe, J. Kleinberg, E. Tardos. Influential Nodes in a Diffusion Model for Social Networks. Lecture Notes in Computer Science, Eds C. Luis, I.Giuseppe et.al, 2005
6. S. Morris. Contagion. Review of Economic Studies 67, 57-78, 2000.
7. L.Backstrom, D. Huttenlocher, J. Kleinbrg, X. Lan, Group Formation in Large Social Networks: Membership, Growth and Evolution, In Proc. KDD 2006.
8. J.L. Iribarren, E. Moro, Impact of Human Activity Patterns on the Dynamics of Information Diffusion, Phys. Rev. Letters, Vol 103, 038702, 2009
9. J. Leskovec, L. Adamic, B. Huberman, The Dynamics of Viral Marketing, EC 2006

Topic 8. Information cascades

Content:
Observational learning. Information cascades.

Recommended reading:
1. Chapter 16 of David Easley and John Kleinberg. "Networks, Crowds, and Markets: Reasoning About a Highly Connected World." Cambridge University Press 2010.

Supplementary reading:

1. Chapter 19 of David Easley and John Kleinberg. "Networks, Crowds, and Markets: Reasoning About a Highly Connected World." Cambridge University Press 2010.
2. A. V. Banerjee. A Simple Model of Herd Behavior. The Quarterly Journal of Economics, Vol. 107, No. 3, pp. 797-817, 1992.
3. S. Bikhchandani, D Hirshleifer and I.Welch. A Theory of Fads, Fashion, Custom, and Cultural Change as Information Cascades. Journal of Political Economy. Vol. 100, pp. 992-1026, 1992.
4. S. Bikhchandani, D Hirshleifer and I.Welch. Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades
5. L. Anderson and C. Halt. Information Cascades in the Laboratory. The American Economic Review, Vol. 87, No. 5 (Dec., 1997), pp. 847-862
6. Pierre Lemieux. Following the Herd . Regulation, Winter 2003-2004

Topic 9. Evolving networks and link prediction

Content:
Network growth. Shrinking diameter. Link reciprocity. Link prediction problem. Supervised learning for prediction.

Recommended reading:

1. L. Backstrom, J. Leskovec. Supervised Random Walks: Predicting and Recommending Links in Social Networks. In Proc. WSDM, 2011
2. D. Liben-Nowell, J. Kleinberg. The Link Prediction Problem for Social Networks. Proc. CIKM, 2003.
3. J. Leskovec, L. Backstrom, R. Kumar, A. Tomkins. Microscopic Evolution of Social Networks. In Proc. KDD 2008.
4. J. Leskovec, J. Kleinberg, C. Faloutsos. Graph Evolution: Densification and Shrinking Diameters. ACM TKDD, 2007.

Supplementary reading:

1. P. D'haeseleer, S. Liang, R. Somogyi. Genetic network inference: from co-expression clustering to reverse engineering. Bioinformatics, vol. 16, 2000.
2. G. Kossinets, D.J. Watts. Empirical Analysis of an Evolving Social Network. Science, 2006.
3. R. Kumar, J. Novak, A. Tomkins. Structure and evolution of online social networks. In Proc. KDD, 2006.

Topic 10. Network visualization Content:
Visual exploration. Graph layouts: radial, force directed, spectral. Matrix visualization.

Recommended reading:

1. "Handbook of Graph Drawing and Visualization", Eds Roberto Tamassia, CRC Press, 2013
2. Brandes, Ulrik, and Dorothea Wagner. "Analysis and visualization of social networks." Graph drawing software. Springer Berlin Heidelberg, 2004. 321-340.
3. Nathalie Henry and Jean-daniel Fekete. "MatrixExplorer: a Dual-Representation System to Explore Social Networks, IEEE Transactions on Visualization and Computer Graphics, 2006. v12 , pp 677-684.

Supplementary reading:

1. V. Batagelj and A. Mrvar. Pajek-analysis and visualization of large networks. Springer, 2003.
2. Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. In Proc ICWSM 2009.
3. Henry, N.; Fekete, J.; McGuffin, M.J., "NodeTrix: a Hybrid Visualization of Social Networks," Visualization and Computer Graphics, IEEE Transactions on , vol.13, no.6, pp.1302,1309,

11. Term Educational Technology

The following educational technologies are used in the study process:

- discussion and analysis of the results of the home task in the group;
- individual education methods, which depend on the progress of each student;
- analysis of skills to formulate common problem in terms of mathematics and solve it;

The course is used to be in format of teleconference lectures.

12. Recommendations for course lecturer

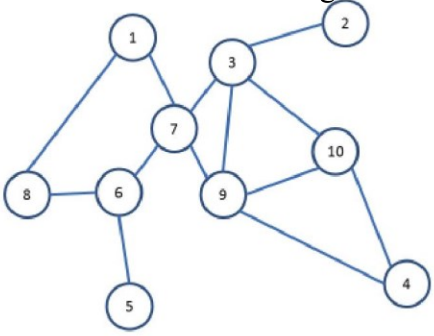
Course lecturer is advised to use interactive learning methods, which allow participation of the majority of students, such as slide presentations, combined with writing materials on board, and usage interactive programming environments for demonstration purposes. The course is intended to be adaptive, but it is normal to differentiate tasks in a group if necessary, and direct fast learners to solve more complicated tasks.

13. Recommendations for students

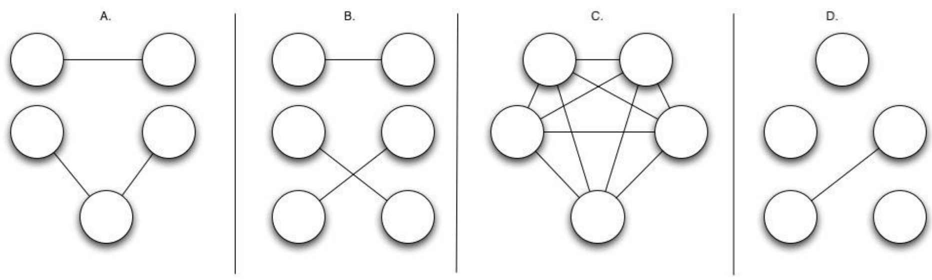
The course is interactive. Lectures are combined with classes. Students are invited to ask questions and actively participate in-group discussions. There will be special office hours for students, which would like to get more precise understanding of each topic. Teaching assistant will also help you. All tutors are ready to answer your questions online by official e-mails that you can find in the “contacts” section**.**

14. Sample final exam questions

1. Some social network has an exponential destribution of node degrees, given by $P(k) = Ce^{-ak}$, where C and a are constants and a is given. Find the fraction of the network nodes that have no more than k_0 neigbours. Use continues approximation $k_{min} = 0$, $k_{max} = \infty$.
2. Consider random walk on the undirected graph given below. Calculate the probability of being on the node #9 in the limit of large time.



3. Describe Zipf’s law. What is the connection between Zipf’s law and power law distributions?
4. Calculate node desgree destributions in the Barabasi-Albert model with uniform probability of attachment P(k) = 1/(m_0 + T)
5. Which of the graphs given below has been most likely generated form Erdos-Renyi model with N=5 and p=0.3 Why? Which one couldn’t be generated by that model?



15. Reading and Materials

15.1 Required Reading

1. Mark Newman. "Networks: An Introduction". Oxford University Press, 2010.
2. David Easley and John Kleinberg. "Networks, Crowds, and Markets: Reasoning About a Highly Connected World." Cambridge University Press 2010.

15.2. Recommended Reading

1. Stanley Wasserman and Katherine Faust. "Social Network Analysis. Methods and Applications." Cambridge University Press, 1994
2. Matthew O. Jackson. "Social and Economic Networks". Princeton University Press, 2010.

15.3. List of review papers

1. R. Albert and A-L. Barabasi. Statistical mechanics of complex networks. Rev. Mod. Phys, Vol. 74, p 47-97, 2002
2. M. E. J. Newman. The Structure and Function of Complex Networks. SIAM Review, Vol. 45, p 167-256, 2003
3. S. Boccaletti et al. Complex networks: Structure and dynamics. Phys. Reports, Vol. 424, p 175-308, 2006
4. S. N. Dorogovtsev and J. F. F. Mendes. Evolution of Networks. Adv. Phys. Vol. 51, N 4, p 1079-1187

15.4. Course telemaintenance

All material of the discipline are posted in informational educational site at NRU HSE portal www.cs.hse.ru/ai, http://www.leonidzhukov.net/hse/2016/networks/, https://github.com/MakarovIA/networks and DataJoy Group.
Students are provided with links to research papers, electronic books, data and software via different cloud services and tools.

16. Equipment

The course requires a laptop, projector, and acoustic systems.
It also requires opportunity to install programming software, such as:

- Python
- Matlab
- R

on student personal computers.
One of core ways is to use the service www.getdatajoy.com or use installed software on

Amazon Web-server.

Lecture materials, course structure and syllabus are prepared by Leonid Zhukov and Ilya Makarov