Government of Russian Federation

Federal State Autonomous Educational Institution of High Professional Education

«National Research University Higher School of Economics»

Faculty of Computer Science
School of Data Analysis and Artificial Intelligence

Syllabus for the course
«Social Networks»

45.04.03 "Fundamental and applied linguistics"

Elective course

Author: Leonid Zhukov, professor, lzhukov@hse.ru

Approved by: Head of Department, Prof. Sergei O. Kuznetsov

Recommended by:

Moscow, 2015

Правительство Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего профессионального образования «Национальный исследовательский университет «Высшая школа экономики»

Факультет компьютерных наук
Департамент анализа данных и искусственного интеллекта

Программа дисциплины

Социальные сети
(“Social Networks” in English)

для направления 45.04.03 Фундаментальная и прикладная лингвистика
подготовки **магистра**

Авторы программы:
Жуков Л.Е., Ph.D., профессор (lzhukov@hse.ru)

Одобрена на заседании кафедры
Анализа данных и искусственного интеллекта «___»____________ 2014 г.
Зав. кафедрой Кузнецов С.О.

Рекомендована секцией УМС
«Прикладная математика и информатика» «___»____________ 2014 г.
Председатель Кузнецов С.О.

Утверждена УС факультета бизнес-информатики «___»_____________2014 г.
Ученый секретарь ______________________________

Москва, 2015

1. Summary

The course "Social Networks" introduces students to the new interdisciplinary field of research. Emerged in sociology, the theory of social networks in recent years, has attracted considerable interest of economists, mathematicians, physicists, experts in data analysis, computer engineers. Initially, researches focused on the study of social networks, i.e. sets of links connecting the social actors in accordance with their interaction. Nowadays, the study of actors' relations includes economic, financial, transport, computer, language and many other networks. The course examines the methods of analyzing the structure of networks, model of their emergence and development, and the processes occurring in networks.

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 the 1st year students of Master Program "Computational Linguistics" in the field of study 45.04.03 «Fundamental and Applied Linguistics».

This syllabus meets the standards required by:

- Educational standards of National Research University Higher School of Economics;
- Educational program of Federal Master's Degree Program for 2015;
- University curriculum of the Master's program in «Computational Linguistics» (45.04.03) for 2015.

3. Learning Objectives

The main objective of the course «Social Networks» – to provide students with the theoretical foundations of the theory of social networks and the development of practical knowledge and skills for network science.

4. Learning outcomes

After completing the study of the discipline « Social Networks » the student should:

- Understand the fundamental principles of social networking
- Know the typical applied problems considered in models of complex networks
- Understand the capabilities and limitations of the existing network analysis methods
- Be able to apply this knowledge to analyze real-world networks.

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

This academic discipline is elective in the educational program "Computational Linguistics" for master 45.04.03 "Fundamental and Applied Linguistics".

For the development of the discipline we assume a basic knowledge in such areas of mathematics and computer science, as "Discrete mathematics", "Probability Theory and Mathematical Statistics", "Informatics and programming", "Algorithms and Data Structures", "Linear algebra" - relevant disciplines included in the program education towards a bachelor 45.04.03 fundamental and applied linguistics.

The following knowledge and competence are needed to study the discipline:

- A good command of the English language, both orally and written.
- A basic knowledge of mathematics
- A basic programming experience

After completing the study of the discipline «Social Networks» 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 network methods in social sciences	Lectures and tutorials.
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 visualize and summarize data, develop mathematical models	Examples covered during the lectures and tutorials. Assignments.
Capability of development of new research methods, change of scientific and industrial profile of self-activities	SC-3	SC-M3	Students obtain necessary knowledge in network science, sufficient to develop new methods in other disciplines.	Assignments, additional material/reading provided.
The ability to construct and solve mathematical models in accordance with the direction of training and specialization	PC-3	IK-M7.2	Student demonstrates the knowledge of basic mathematical models used in the analysis of social networks	Analysis of features of mathematical models, the solution of problems in the use of social networking models
Ability to use in research and applied activities of modern programming languages and data manipulation	PC-4	IK-M7.5	Student is confident the program in high-level language for computing the application based on the social network model	Homework oriented on software implementation models; analysis of experimental data.

Competence	Code	Code (UC)	Descriptors (indicators of achievement of the result)	Educative forms and methods aimed at generation and development of the competence
languages, operating systems, software packages, etc.				
Ability to publicly present the results of professional activity (including the use of information technologies)	PC-5	IK-M2.5	Student is able to understand, analyze and present the report in the form of compressed material of scientific articles in the field of social network analysis	Preparation of independent course projects on social networks with following oral defense

6. Schedule of the course “Social Networks”

Two pairs consist of 2 academic hour for lecture followed by 2 academic hour for computer exercises/labs after lecture. Additional office hours for lectures’ content are provided.

№	Название темы	Всего часов по дисциплине	Аудиторные часы		Самостоятельная работа
			Лекции	Семинары занятия	
1	Introduction to network science	21	3	3	15
2	Node and link analysis	21	3	3	15
3	Network structure and communities	21	3	3	15
4	Network visualization	21	3	3	15
5	Linguistic networks: word co-occurrence, semantic and syntactic networks	30	4	4	22
	Total:	114	16	16	82

7. Grading

Type of grading	**Type of work**	Параметры
Current	Mid Term Exam	Written work for 80 minutes
Current	Homework	2 weeks
Final	Exam	Written work for 80 minutes

8. Assessment

In the current and final control student must demonstrate knowledge of the basic concepts of the discipline passed.

The current control includes written tests, consisting of multiple issues and challenges on the passed material and homework on the use of the models and analysis of experimental data.

The final control is carried out in the form of written work, including several issues and challenges on the Topics of discipline.

The grade formula:

The current control will consist of midterm exam and homework, worth 50% each one

Current=0.5*(Homework)+0.5*(MidTerm).

The exam will consist of problems, worth 50% of the final mark.

Final course mark *is obtained from the following formula*: Final=0.5*(Current)+0.5*(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 Grading Scale	**Five-point 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	

9. Course Description

Topic 1. Complex networks.

Introduction to the theory of complex systems. Basic concepts in the theory of networks. Properties and network analysis metrics. The power-law distribution. Scale-invariant network (scale-free networks). Random graphs. Pareto distribution, normalization, moments Act Tsipfa.Graf rank-frequency diameter and clustering coefficients.

Recommended 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.

Supplementary 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.

Topic 2. Nodes metrics and link analysis

Metrics and central nodes / Centrality metrics. The concepts of centrality and prestige. Model graphs. Degree centrality, closeness centrality, betweenness centrality, status / rank prestige (eigenvector centrality). Central network (sentralization). Analysis of bonds. PageRank algorithm. Stochastic matrices. Hubs and Authorities. HITS algorithm.

Recommended 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.

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. Networks in theoretical linguistics

Networks and semantics. Networks and syntax. Networks and morphology. Networks and phonology. Networks and applied linguistics.

Recommended Reading

"Structure Discovery in Natural Language". Chris Biemann, Springer, 2012.

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.

Paper Review and in-course reading

Introductory Articles

- Scale Free Networks, Scientific American, A.L Barabasi

- The physics of networks, Physics Today, Mark Newman

Reviews

- Statistical mechanics of complex networks, R. Albert and A-L. Barabasi
- The Structure and Function of Complex Networks, M. E. J. Newman
- Complex networks: Structure and dynamics, S. Boccaletti et al.

Research papers on linguistic networks

- Language Networks: Their Structure, Function and Evolution Ricard V. Sole, Bernat Corominas Murtra, Sergi Valverde, Luc Steels
- Semantic Networks: Structure and Dynamics Javier Borge-Holthoefer and Alex Arenas
- Quantifying Semantics Using Complex Network Analysis Chris Biemann, Stefanie Roos, Karsten Weihe.
- The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth Mark Steyversa, Joshua B. Tenenbaumb
- The small world of human language. Ramon Ferrer i Cancho and Ricard V. Sole
- Complex network analysis of literary and scientific texts. Iwona Grabska-Gradzinska, Andrzej Kulig, Jaroslaw Kwapien, Stanislaw Drozdz
- The Structure and Dynamics of Linguistic Networks Monojit Choudhury, Animesh Mukherjee.
- Networks in Cognitive Science A. Baronchelli, R Ferrer-i-Cancho, R. Pastor-Satorras, N. Chater, M. Christiansen

10. Term Educational Technology

The following educational technologies are used in the study process:

- discussion and analysis of the results during the computer exercises;
- regular assignments to test the progress of the student;
- teleconference lectures
- office hours and classes with tutor and teaching assistants
- paper reviews
- oral defense of HW assignments
- tutorship

11. Evaluation tools for current and final control

Homework samples

1. Build a degree distribution of nodes in the fragment "internet routing system". (Http://www.routeviews.org/). Show that the distribution follows a power law. Rate exponent value of the slope of the curve. Construct a cumulative distribution function, find the value of the exponent from the slope of the curve. Calculate the exponent using the maximum likelihood method. Compare the results.

2. Calculate a correlation matrix (Pearson correlation) structural equivalence of nodes in the networks "karate_club" and "dolphins" and visualize it pcolor team. 2) Calculate the amount of assortative mixing powers of nodes (assortativity coefficient) in the networks of "Princeton", "Georgetown", "internet autonomous", "political blogs".

3. Implement the SIS/SIR epidemic models in networks. Investigate the behavior of the models on the following networks: 1) regular two-dimensional grid 2) two-dimensional model of small world 3) random graph 4) a model of the preferred joining BA 5) this network. For each model / build network averaged dependence of infection (% infected hosts) from time to time for a fixed choice of model parameters.

Midterm exam and exam problem samples

1. What is the distribution of (the law) Zipf? As it is connected with the power distribution (power law)?
2. Matrix of influence in social influence DeGroot model given matrix

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

   Show the diagram of influence and computed the influence for each of the participants. If reaching a consensus is possible, calculate the limiting beliefs with initial beliefs f = (1/4, 1/2, 1).

12. Questions for assessing the quality of study

Topic 1.
1. What systems and networks are called complex (scale invariant)
2. Give examples of complex networks

Topic 2.
1. Give examples of the Pareto distribution
2. How to build a graph of rank-frequency?

Topic 3.
1. Tell about the model of the preferred connection
2. Explain the concept of "small world"

Topic 4.
1. Describe the different types of central nodes
2. Describe the PageRank algorithm

Topic 5.
1. Define the network community
2. Explain the concept of modularity

Topic 6.
1. Describe the process of diffusion on the networks
2. What is the difference SIS/SIR models?

Topic 7.
1. Share threshold models of decision-making
2. Describe the algorithm for finding the most influential nodes

Topic 8.
1. Tell about the models learning networks
2. What are the conditions for achieving a consensus?

Topic 9.
1. What information is called a cascade?
2. What are the conditions necessary for the emergence of stage?

Topic 10.
1. What is the model of Schelling?
2. How to change the findings in the network model Schelling case?

13. Reading and Materials

Core book – short conspectus based on the following sources:

1. Mark Newman. "Networks: An Introduction". Oxford University Press, 2010.
2. Matthew O. Jackson."Social and Economic Networks". Princeton University Press, 2010.
3. David Easley and John Kleinberg. "Networks, Crowds, and Markets: Reasoning About a Highly Connected World." Cambridge University Press 2010.
4. Stanley Wasserman and Katherine Faust. "Social Network Analysis. Methods and Applications." Cambridge University Press, 1994

Supplementary Reading

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. N. Dorogovtsev and J. F. F. Mendes. Evolution of Networks. Adv. Phys. Vol. 51, N 4, p 1079-1187
4. A.L Barabasi. Scale Free Networks. Scientific American, p 50-59, 2003
5. Mark Newman The physics of networks. Physics Today,2008
6. M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. Contemporary Physics 46(5), 323-351, 2005
7. Clauset, C.R. Shalizi, M.E.J. Newman. Power-law distributions in empirical data. SIAM Review 51(4), 661-703, 2009
8. M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. Internet Mathematics, vol 1, No. 2, pp. 226-251, 2004

9. 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.
10. P. Erdos and A. Renyi. On random graphs I. Publ. Math. Debrecen, 1959.
11. P. Erdos and A. Renyi. On the evolution of random graphs. Magyar Tud. Akad. Mat. Kutato Int. Koezl., 1960.
12. Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. . Nature 393:440-42, 1998.
13. AL Barabasi and R. Albert. Emergence of Scaling in Random Networks. Science, 286, 1999.
14. Linton C. Freeman. Centrality in Social Networks. Conceptual Clarification. Social Networks, Vol 1, pp 215-239, 1978
15. Phillip Bonacich. Power and Centrality: A Family of Measures. American journal of sociology, Vol.92, pp 1170-1182, 1987.
16. S. Brin, L. Page. The PageRank Citation Ranknig: Bringing Order to the Web.
17. John M. Kleinberg. Authoritative Sources in a Hyperlinked Environment. Proc. 9th ACM-SIAM Symposium on Discrete Algorithms, 1998.
18. M. Newman. Mixing patterns in networks. Phys. Rev. E, Vol. 67, p 026126, 2003
19. M.E.J. Newman, M. Girvan. Finding and evaluating community structure in networks. Phys. Rev. E 69, 026113, 2004.
20. M.E.J. Newman. Modularity and community structure in networks. PNAS Vol. 103, N 23, pp 8577-8582, 2006
21. D.R. Karger. Global min-cuts in RNC, and other ramifications of a simple min-cut algorithm. Proceedings SODA '93, p. 21-30, 1993
22. Abou-rjeili, G. Karypis. Multilevel algorithms for partitioning power-law graphs. In Proceedings IPDPS '06, p 10, 2006
23. 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.
24. 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
25. R. Andersen, F. Chung, K. Lang. Local graph partitioning using pagerank vectors. In Proc. FOCS, 2006.
26. S. E. Schaeffer. Graph clustering. Comp. Sci. Rev., Vol. 1, p 27-64, 2007
27. S. Fortunato. Community detection in graphs . Physics Reports, Vol. 486, pp. 75-174, 2010
28. V. Batagelj, M. Zaversnik. An O(m) Algorithms for Cores Decomposition of Networks. 2003
29. 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
30. The strength of weak ties. M. Granovetter. American Journal of Sociology, 78(6):1360-1380, 1973.
31. H.W. Hethcote. The Mathematics of Infections Diseases. SIAM Review, Vol. 42, No. 4, pp. 599-653
32. Matt. J. Keeling and Ken.T.D. Eames. Networks and Epidemics models Journal R. Soc. Interface, Vol 2, pp 295-307, 2005
33. D. Kempe, J. Kleinberg, E. Tardos. Maximizing the Spread of Influence through a Social Network. In Proc. KDD 2003.
34. 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
35. M. Richardson, P. Domingos. Mining Knowledge-Sharing Sites for Viral Marketing. In Proc. KDD, 2002.
36. M. Richardson, P. Domingos. Mining the Network Value of Customers. In Proc. KDD, 2001.

37. M.H. DeGroot. Reaching a Consensus. Journal of the American Statistical Association, 1974, Vol 69, No 345
38. Roger L. Berger. A Necessary and Sufficient Condition for Reaching a Consensus Using DeGroot's Method. Journal of the American Statistical Association, Vol. 76, No 374, 1981, pp 415-418
39. 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.
40. D. Watts. A simple model of global cascades on random networks. Proc. Natl. Acad. Sci., vol. 99 no. 9, 5766-5771, 2002.
41. H. P. Young. The Diffusion of Innovations in Social Networks. Santa Fe Institute Working Paper 02-04-018.
42. S. Bikhchandani, D Hirshleifer and I.Welch. Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades
43. L. Anderson and C. Halt. Information Cascades in the Laboratory. The American Economic Review, Vol. 87, No. 5 (Dec., 1997), pp. 847-862
44. Pierre Lemieux. Following the Herd . Regulation, Winter 2003-2004
45. S. Morris. Contagion. Review of Economic Studies 67, 57-78, 2000.
46. A.V. Banerjee. A Simple Model of Herd Behavior. The Quarterly Journal of Economics, Vol. 107, No. 3, pp. 797-817, 1992.
47. Thomas C. Schelling Dynamic Models of Segregation , Journal of Mathematical Sociology, Vol. 1, pp 143-186, 1971.
48. Arnaud Banos Network effects in Schellin's model of segregation: new evidences from agent-based simulations. Environment and Planning B: Planning and Design Vol.39, no. 2, pp. 393-405, 2012.
49. Giorgio Gagiolo, Marco Valente, Nicolaas Vriend Segregation in netwroks. Journal of Econ. Behav. & Organization, Vol. 64, pp 316-336, 2007.
50. Mark S. Granovetter. Threshold Models of Collective Behavior. American Journal of Sociology Vol. 83, No. 6, pp. 1420-1443, 1978.

14. Equipment

The following software tools are recommended for the students:

- Computations: Matlab, Octave, R, Python
- Graph Visualization: Gephi, yEd
- Matlab libraries:
  - Graph algorithms: MatlabBGL
  - Graph layout: GraphViz., gpaphviz4matlab
- Python libraries:
  - Linear algebra algorithms: Numpy, SciPy
  - Graph algorithms and visualizaton: NetworkX, iGraph
- R libraries:
  - Graph algorithms: igraph, statnet

Lecture materials, course structure and the syllabus are prepared by Leonid E. Zhukov.