

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО
Директор физтех-школы
прикладной математики
и информатики
А.М. Райгородский

Рабочая программа дисциплины (модуля)

Дисциплина:	Анализ сетевых данных
Направление:	Прикладные математика и физика
Магистерская программа:	Интеллектуальный анализ данных Физтех-школа прикладной математики и информатики Кафедра проблем передачи информации и анализа данных
Курс:	1
Квалификация:	Магистр

Семестр, формы промежуточной аттестации: 1 (Осенний) – Дифференцированный зачёт

Аудиторных часов: 30 всего, в том числе:

лекции: 30 час.

практические (семинарские) занятия: 0 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 15 час. всего, в том числе:

задания, курсовые работы: 0 час.

Подготовка к экзамену: 0 час.

Всего часов: 45, всего зач.ед.: 1

Программу составили: **М.Е. Панов, кандидат физико-математических наук**
А.И. Курмуков, кандидат компьютерных наук

Аннотация

Учебный курс ориентирован на изучение методов и технологий, используемых при сборе и анализе данных в сети Интернет, а также на получение практических навыков разработки программных решений для автоматизации данных задач. В рамках курса изучаются подходы к извлечению данных из неструктурированных источников (корпуса текстов, web-страницы, социальные сети), вопросы обработки мультимедийной информации и естественного языка, тенденции развития Web-технологий.

Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 1 (Осенний)

1. Введение в теорию графов

Вершины и ребра графа. Путь. Цикл. Эйлеров граф. Гамильтонов граф. Связность графа, компоненты связности. Ориентированные графы. Ориентированный ациклический граф. Двудольный граф. Взвешенные графы. Полный граф. Деревья. Алгоритмы поиска в графах. Примеры реальных сетей. Граф социальной сети. Эгоцентрические графы. Графы цитирования. Информационные сети. Биологические сети.

2. Модели порождения сетей

Модель Эрдёша-Рени. Модели “малого мира” и “предпочтительного присоединения”. Модель Ваттса-Строгаца. Модель Барабаси-Альберта. Стохастическая блочная модель. Графы Кронекера.

3. Алгоритмы анализа сетей

Характеристики сетей. Степень вершины. Степенной закон распределения вершин. Масштабно-инвариантные сети. Диаметр графа. Средняя длина пути. Меры центральности. Локальные и глобальные характеристика графа. Спектральные меры вершин графа. Пейдж ранк. Алгоритм HITS. Работа с сетями с помеченными вершинами. Коэффициент ассортативности.

4. Динамические сети и модели распространения информации

Модели распространения эпидемии: SI, SIS, SIR. Выделение ключевых вершин. Модели влияния. Подходы к максимизации влияния в сетях. Распространение информации в реальных сетях.

5. Кластеризация вершин графа. Понятие модулярности.

Разбиения вершин графа на сообщества. Модулярность как характеристика качества кластеризации вершин. Способы максимизации модулярности. Дивизимные и агломеративные способы кластеризации вершин. Пересекающиеся сообщества.

6. Машинное обучение на графах

Предсказание наличия ребер в “растущих” графах. Классификация вершин графа на основе структуры внутренних связей. Использование “не сетевой” информации для улучшения моделей предсказания ребер и классификации вершин.

Перечень основной и дополнительной литературы, необходимой для освоения дисциплины (модуля)

Основная литература

1. Newman M.E.J. Networks - An introduction, Oxford Univ Press, 2010.
2. Wasserman S., Faust K. Social Network Analysis: Methods and Applications, Cambridge Univ Press, 1994.

Дополнительная литература

1. Erdos P., Renyi A. *On Random Graphs I*. Publ. Math. Debrecen, 1959.
2. Erdos P., Renyi A. *On the evolution of random graphs*, 1960.
3. Magyar Tud. Akad. Mat. Kutato Int. Koezl., 1960.
4. Newman M.E.J. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 2005. 46(5). pp. 323-351.
5. Linton C. Freeman. Centrality in Social Networks. Conceptual Clarification. *Social Networks*, 1978. Vol. 1. pp. 215-239.
6. Fortunato S. Community detection in graphs. *Physics Reports*, 2010. Vol. 486. pp. 75-174.
7. Milgram S. *The small world problem*. *Psychology Today* 1, 1967.
8. Newman M.E.J. Models of the Small World: A Review. *J. Stat. Physics*, 2000.
9. Broido Anna D., Clauset A. Scale-free networks are rare, 2018.
10. Page L. The PageRank Citation Ranking: Bringing Order to the Web, 1998.
11. Zhu X., Ghahramani Z. Learning from Labeled and Unlabeled Data with Label Propagation, 2002.
12. Blondel V.D., Guillaume J.-L., Lambiotte R., Lefebvre E. Fast unfolding of communities in large networks, 2008.