

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

УТВЕРЖДЕНО
Проректор по учебной работе

А.А. Воронов

	Рабочая программа дисциплины (модуля)
по дисциплине:	Анализ сетевых данных
по направлению:	Информатика и вычислительная техника
профиль подготовки:	Прикладная математика и информатика Физтех-школа Прикладной Математики и Информатики кафедра проблем передачи информации и анализа данных
курс:	1
квалификация:	магистр

Семестр, формы промежуточной аттестации: 1 (осенний) - Дифференцированный зачет

Аудиторных часов: 30 всего, в том числе:

лекции: 15 час.

семинары: 15 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 15 час.

Всего часов: 45, всего зач. ед.: 1

Программу составил: М.Е. Панов, канд. физ.-мат. наук

Программа обсуждена на заседании кафедры проблем передачи информации и анализа данных 04.06.2020

Аннотация

Целью освоения дисциплины является знакомство с основными понятиями и методами анализа сетей; практические навыки анализа и моделирования сетей.

1. Цели и задачи

Цель дисциплины

Дать представление о современном состоянии теории сложных сетей и ее использовании в анализе данных.

Задачи дисциплины

- изучение моделей сложных сетей и их теоретического обоснования;
- практическое применение моделей сложных сетей в задачах анализа данных.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-3 Способен выбирать и (или) разрабатывать подходы к решению типовых и новых задач в области информатики и вычислительной техники, учитывая особенности и ограничения различных методов решения	ОПК-3.5 Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий
	ОПК-3.2 Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем
ПК-2 Понимает и способен применить в научно-исследовательской и прикладной деятельности основные законы естествознания, современный математический аппарат и алгоритмы, современные информационно-коммуникационные технологии	ПК-2.2 Умеет применять полученные знания в области фундаментальных научных основ теории информации и решать стандартные задачи в собственной научно-исследовательской деятельности

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- базовые модели сложных сетей;
- подходы к анализу сетевых данных;
- примеры использования сложных сетей в прикладных задачах.

уметь:

- составлять модели сетей из неструктурированных данных;
- определять класс реальной сети и подходящую для нее теоретическую модель;
- проводить анализ сложных сетей;
- оценивать закон распределения степеней вершин.

владеть:

- навыком освоения большого объема информации;
- навыками постановки научно-исследовательских задач и навыками самостоятельной работы.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.
--	---

№	Тема (раздел) дисциплины	Лекции	Семинары	Лаборат. работы	Самост. работа
1	Введение в теорию графов	2	2		2
2	Модели порождения сетей	2	2		2
3	Алгоритмы анализа сетей	2	2		2
4	Динамические сети и модели распространения информации	3	3		2
5	Кластеризация вершин графа. Понятие модулярности	3	3		2
6	Машинное обучение на графах	3	3		5
Итого часов		15	15		15
Подготовка к экзамену		0 час.			
Общая трудоёмкость		45 час., 1 зач.ед.			

4.2. 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. 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.

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Учебная аудитория, оснащенная мультимедийным оборудованием (проектор или плазменная панель), доской.

6. Перечень рекомендуемой литературы

Основная литература

1. Основы сетевой безопасности: криптографические алгоритмы и протоколы взаимодействия [Текст] : курс лекций : учеб. пособие для вузов / О. Р. Лапоница ; под ред. В. А. Сухомлина .— М. : Интернет-Ун-т Информ. Технологий, 2005 .— 608 с.
фонд базовой кафедры
2. Newman M.E.J. Networks - An introduction, Oxford Univ Press, 2010.
3. Wasserman S., Faust K. Social Network Analysis: Methods and Applications, Cambridge Univ Press, 1994.

Дополнительная литература

1. Графы. Алгоритмы на языке C [Текст] / В. В. Прут ; М-во образования и науки РФ, Моск. физ.-техн. ин-т (гос. ун-т) - М. МФТИ, 2017
2. Графы и их применение. Комбинаторные алгоритмы для программистов [Текст] / Н. И. Костюкова - М. Интернет-Ун-т Информ. Технологий : БИНОМ. Лаб. знаний, 2007
фонд базовой кафедры
3. Erdos P., Renyi A. On Random Graphs I. Publ. Math. Debrecen, 1959.
4. Erdos P., Renyi A. On the evolution of random graphs, 1960.
5. Magyar Tud. Akad. Mat. Kutato Int. Koezl., 1960.
6. Newman M.E.J. Power laws, Pareto distributions and Zipf's law. Contemporary Physics, 2005. 46(5). pp. 323-351.
7. Linton C. Freeman. Centrality in Social Networks. Conceptual Clarification. Social Networks, 1978. Vol. 1. pp. 215-239.
8. Fortunato S. Community detection in graphs. Physics Reports, 2010. Vol. 486. pp. 75-174.
9. Milgram S. The small world problem. Psychology Today 1, 1967.
10. Newman M.E.J. Models of the Small World: A Review. J. Stat. Physics, 2000.
11. Broido Anna D., Clauset A. Scale-free networks are rare, 2018.
12. Page L. The PageRank Citation Ranking: Bringing Order to the Web, 1998.
13. Zhu X., Ghahramani Z. Learning from Labeled and Unlabeled Data with Label Propagation, 2002.
14. Blondel V.D., Guillaume J.-L., Lambiotte R., Lefebvre E. Fast unfolding of communities in large networks, 2008.

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>).

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

не требуется

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Студент, изучающий дисциплину, должен, с одной стороны, овладеть общими понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике.

В результате изучения дисциплины студент должен знать основные определения, понятия, аксиомы, методы доказательств.

Успешное освоение курса требует напряженной самостоятельной работы студента. В программе курса отведено минимально необходимое время для работы студента над темой. Самостоятельная работа включает в себя:

- чтение и конспектирование рекомендованной литературы;
- проработку учебного материала (по конспектам занятий, учебной и научной литературе), подготовку ответов на вопросы, предназначенные для самостоятельного изучения, доказательство отдельных утверждений, свойств, решение задач;
- подготовка к дифференцированному зачёту.

Руководство и контроль за самостоятельной работой студента осуществляется в форме индивидуальных консультаций.

Важно добиться понимания изучаемого материала, а не механического его запоминания. При затруднении изучения отдельных тем, вопросов следует обращаться за консультациями к лектору.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Информатика и вычислительная техника
профиль подготовки:	Прикладная математика и информатика Физтех-школа Прикладной Математики и Информатики кафедра проблем передачи информации и анализа данных
курс:	<u>1</u>
квалификация:	магистр

Семестр, формы промежуточной аттестации: 1 (осенний) - Дифференцированный зачет

Разработчик: М.Е. Панов, канд. физ.-мат. наук

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-3 Способен выбирать и (или) разрабатывать подходы к решению типовых и новых задач в области информатики и вычислительной техники, учитывая особенности и ограничения различных методов решения	ОПК-3.5 Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий
	ОПК-3.2 Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем
ПК-2 Понимает и способен применить в научно-исследовательской и прикладной деятельности основные законы естествознания, современный математический аппарат и алгоритмы, современные информационно-коммуникационные технологии	ПК-2.2 Умеет применять полученные знания в области фундаментальных научных основ теории информации и решать стандартные задачи в собственной научно-исследовательской деятельности

2. Показатели оценивания компетенций

В результате изучения дисциплины «Анализ сетевых данных» обучающийся должен:

знать:

- базовые модели сложных сетей;
- подходы к анализу сетевых данных;
- примеры использования сложных сетей в прикладных задачах.

уметь:

- составлять модели сетей из неструктурированных данных;
- определять класс реальной сети и подходящую для нее теоретическую модель;
- проводить анализ сложных сетей;
- оценивать закон распределения степеней вершин.

владеть:

- навыком освоения большого объема информации;
- навыками постановки научно-исследовательских задач и навыками самостоятельной работы.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

С целью контроля освоения обучающимися учебного материала проводится устный опрос в начале занятия по теме прошлой лекции или в конце занятия по пройденной теме.

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

1. Генерация различных моделей случайных графов. Степенной закон распределения вершин. Подсчет коэффициента кластеризации. Поиск в графе.
2. Реализация подсчета различных метрик центральности. Сравнение характеристик реальных графов.
3. Алгоритмы поиска сообществ в сетях.
4. Анализ графа эгоцентрической сети графа социальной сети.
5. Предсказание связей в графе цитирования.
6. Анализ распространения информации в графе социальной сети.

Критерии оценивания

Оценка отлично 10 баллов - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины, проявляющему интерес к данной предметной области, продемонстрировавшему умение уверенно и творчески применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично 9 баллов - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично 8 баллов - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений, с некоторыми недочетами.

Оценка хорошо 7 баллов - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но недостаточно грамотно обосновывает полученные результаты.

Оценка хорошо 6 баллов - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности.

Оценка хорошо 5 баллов - выставляется студенту, если он в основном знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач достаточно большое количество неточностей.

Оценка удовлетворительно 4 балла - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он освоил основные разделы учебной программы, необходимые для дальнейшего обучения, и может применять полученные знания по образцу в стандартной ситуации.

Оценка удовлетворительно 3 балла - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, допускающему ошибки в формулировках базовых понятий, нарушения логической последовательности в изложении программного материала, слабо владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и с трудом применяет полученные знания даже в стандартной ситуации.

Оценка неудовлетворительно 2 балла - выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных принципов и не умеет использовать полученные знания при решении типовых задач.

Оценка неудовлетворительно 1 балл - выставляется студенту, который не знает основного содержания учебной программы дисциплины, допускает грубейшие ошибки в формулировках базовых понятий дисциплины и вообще не имеет навыков решения типовых практических задач.

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Дифференцированный зачёт складывается из выполнения 4-х домашних заданий, итогового проекта и аудиторной активности.