

**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**

	<b>Рабочая программа дисциплины (модуля)</b>
<b>по дисциплине:</b>	Структурный анализ и визуализация сетей
<b>по направлению:</b>	Информатика и вычислительная техника
<b>профиль подготовки:</b>	Прикладная математика и информатика Физтех-школа Прикладной Математики и Информатики кафедра алгоритмов и технологий программирования
<b>курс:</b>	2
<b>квалификация:</b>	магистр

Семестр, формы промежуточной аттестации: 3 (осенний) - Экзамен

Аудиторных часов: 45 всего, в том числе:

лекции: 30 час.

семинары: 15 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 60 час.

Подготовка к экзамену: 30 час.

Всего часов: 135, всего зач. ед.: 3

Количество контрольных работ, заданий: 2

Программу составил: И.А. Макаров, старший преподаватель

Программа обсуждена на заседании кафедры алгоритмов и технологий программирования 13.05.2024

## Аннотация

Курс «Структурный анализ и визуализация сетей» вводит студентов в новую и активно развивающуюся междисциплинарную область наук о сетевых структурах. Начиная с исследований социальных сетей социологами, данное направление привлекло внимание физиков, ученых в области компьютерных наук, экономистов, вычислительных биологов, лингвистов и других, и стало по-настоящему междисциплинарной областью изучения. Несмотря на разнообразие процессов, которые образуют сети, а также объекты и отношения, которые служат узлами и ребрами в этих сетях, все сети обладают общими статистическими и структурными свойствами. Взаимодействие между порядком и беспорядком создает сложные сетевые структуры, которые находятся в центре внимания исследования. В ходе курса мы рассмотрим методы статистического и структурного анализа сетей, модели формирования и эволюции сети, и развития процессов в сети. Особое внимание будет уделено практическому анализу и визуализации реальных сетей с использованием доступных программных средств, и современных языков и библиотек программирования.

### 1. Цели и задачи

#### Цель дисциплины

- ознакомление учащихся с новой быстро развивающейся областью сетевых наук и предоставление практического опыта в анализе сетевых данных реального мира.

#### Задачи дисциплины

- приобретение студентами навыков алгоритмического анализа сетевых графов.

### 2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-2 Способен управлять проектом на всех этапах его жизненного цикла	УК-2.2 Способен прогнозировать результат деятельности и планировать последовательность шагов для достижения данного результата. Формирует план-график реализации проекта в целом и план контроля его выполнения
	УК-2.3 Способен организовать и координировать работу участников проекта, обеспечивать работу команды необходимыми ресурсами
ОПК-2 Имеет представление об актуальных проблемах науки и техники в области информатики и вычислительной техники, способен на научном языке формулировать профессиональные задачи	ОПК-2.1 Имеет представление о современном состоянии исследований в рамках тематической области своей профессиональной деятельности
	ОПК-2.2 Способен оценивать актуальность исследований в области информатики и вычислительной техники и их практическую значимость
	ОПК-2.3 Владеет профессиональной терминологией, используемой в современной научно-технической литературе, обладает навыками устного и письменного изложения результатов научной деятельности в рамках профессиональной коммуникации
ПК-2 Понимает и способен применить в научно-исследовательской и прикладной деятельности основные законы естествознания, современный математический аппарат и алгоритмы	ПК-2.1 Знает основы научно-исследовательской деятельности в области информационных технологий, владеет знанием основ философии и методологии науки; знанием методов научных исследований и навыками их проведения
	ПК-2.2 Умеет применять полученные знания в области фундаментальных научных основ теории информации и решать стандартные задачи в собственной научно-исследовательской деятельности

математический аппарат и алгоритмы, современные информационно-коммуникационные технологии	ПК-2.3 Имеет практический опыт научно-исследовательской деятельности в области информационно-коммуникационных технологий
	ПК-2.4 Владеет методами и алгоритмами решения задач цифровой обработки сигналов, использования сети Интернет, аннотирования, реферирования, библиографического поиска, опыт работы с научными источниками

### 3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны знать:

- знать основные понятия и терминологию, используемые в науке о сетях;
- знать основополагающие принципы сетевой структуры и эволюции.

уметь:

- уметь разрабатывать математические модели сетевых процессов;
- уметь анализировать сетевые данные реального мира.

владеть:

- владеть основополагающими принципами сетевой структуры и эволюции.

### 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

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Введение в науку о сетях	3	1		4
2	Спенной закон	3	1		6
3	Модели формирования сетей	3	1		6
4	Анализ структуры сетей, узлов и связей	3	1		8
5	Сообщества в сетях	3	1		6
6	Диффузия и эпидемии в сетях	3	2		6
7	Распространение влияния	3	2		6
8	Информационные каскады	3	2		6
9	Развитие сетей и предсказание связей	3	2		4
10	Визуализация сетей	3	2		8
Итого часов		30	15		60
Подготовка к экзамену		30 час.			
Общая трудоёмкость		135 час., 3 зач.ед.			

#### 4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 3 (Осенний)

##### 1. Введение в науку о сетях

Введение в новую науку о сетях. Теория сложных сетей. Основные свойства и метрики сетей. Примеры сетей.

## 2. Спенной закон

Распределение степенного закона. Scale-free сети. Распределение Парето, нормализация, моменты. Закон Зипфа. График частот.

## 3. Модели формирования сетей

Модель случайного графа Эрдоса-Рени. Распределения Пуассона и Бернулли. Распределение степеней узлов. Фазовый переход, гигантский связанный компонент. Диаметр и коэффициент кластера. Модель конфигурации. Модель Барабаси-Альберта. Предпочтительное прикрепление. Уравнение в непрерывном приближении. Временная эволюция степеней узлов. Распределение степени узла. Средняя длина пути и коэффициент кластеризации. Модель малого мира. Модель Watts-Strogats. Переход от регулярного к случайному графу, изменения коэффициента кластеризации и средней длины пути.

## 4. Анализ структуры сетей, узлов и связей

Метрики центральности узлов, степенная центральность, центральность близости, betweenness центральность, eigenvector центральность. Метрики, основанные на структуре графа. PageRank, стохастическая метрика и теорема Перрона-Фробениуса. Итерации мощности. Центры и авторитеты. Алгоритм HITS. Ранговое расстояние Кендалл-Тау. Метрики структурной эквивалентности. Расстояния Евклида и Хемминга. Коэффициент корреляции. Косинусное сходства. Ассортативное смешивание и гомофилия. Модульность. Смешивание по степеням.

## 5. Сообщества в сетях

Сообщества в сетях. Плотность графа. Разделение графа. Показатели Min-cut, quotient и normalized cuts. Дивизионные и агломерационные алгоритмы. Повторяющееся деление пополам. Корреляционная матрица. Кластеризация. Edge Betweenness. Алгоритм Ньюмана-Гирвина. Спектральные методы. Алгоритм максимизации модульности (Newman). Аппроксимационные алгоритмы. Рандомизированный минимальный разрез (алгоритм Каргеса). Вероятность нахождения минимального разреза. Многоуровневая парадигма. Многоуровневый алгоритм для графов со степенным законом. Локальная кластеризация. Проводимость. Алгоритмы Nibble. Graph motifs, k-cores, перепись диад и триад.

## 6. Диффузия и эпидемии в сетях

Физическая диффузия. Уравнение диффузии. Диффузия в сетях. Дискретный оператор Лапласа, матрица Лапласа. Решение уравнения диффузии. Случайные блуждания по графу. Эпидемические модели SI, SIS, SIR. Решение дифференциального уравнения. Ограничивающие случаи. Моделирование распространения инфекции.

## 7. Распространение влияния

Социальная диффузия. Пороговая модель коллективного поведения Грановеттера. Наиболее влиятельные узлы в сети. Каскады в сетях. Базовая каскадная модель. Каскадная модель.

## 8. Информационные каскады

Наблюдательное обучение. Информационные каскады.

## 9. Развитие сетей и предсказание связей

Рост сети. Уменьшение диаметра. Взаимодействие связей. Задача предсказания связей. Обучение с учителем для предсказания.

## 10. Визуализация сетей

Визуальное исследование. Графические схемы: радиальные, силовые, спектральные. Матричная визуализация.

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

Учебная аудитория, оснащенная компьютером и мультимедийным оборудованием (проектор, звуковая система).

## 6. Перечень рекомендуемой литературы

### Основная литература

1. Изучаем Python, программирование игр, визуализация данных, веб-приложения/Эрик Мэтиз, Python Crash Course,-Санкт-Петербург, Питер, 2020
2. Изучаем Python: программирование игр, визуализация данных, веб-приложения, [руководство] / Э. Мэтиз. — Санкт-Петербург, Питер, 2020.— URL: <https://ibooks.ru/bookshelf/371712/reading> (дата обращения: 24.11.2020). - Полный текст (Режим доступа : из сети МФТИ / Удаленный доступ)

### Дополнительная литература

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

Подборка интересных Jupyter-ноутбуков:

<https://github.com/jupyter/jupyter/wiki/A-gallery-of-interesting-Jupyter-Notebooks>

База сетевых датасетов из Стенфорда: <https://snap.stanford.edu/data/>

Канал курса в Youtube:

<https://www.youtube.com/playlist?list=PLriUvS7IljvkBLqU4nPOZtAkp7rgpxjg1>

## 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)

Успешное освоение дисциплины требует:

- посещения студентом всех видов аудиторных занятий;
- ведения конспекта в ходе лекционных занятий;
- качественной самостоятельной подготовки к практическим занятиям, активной работы на них;
- активной самостоятельной и аудиторной работы студента;
- своевременной сдачи преподавателю заданий по аудиторным видам работ.

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

<b>по направлению:</b>	Информатика и вычислительная техника
<b>профиль подготовки:</b>	Прикладная математика и информатика Физтех-школа Прикладной Математики и Информатики кафедра алгоритмов и технологий программирования
<b>курс:</b>	<u>2</u>
<b>квалификация:</b>	магистр
Семестр, формы промежуточной аттестации: 3 (осенний) - Экзамен	
<b>Разработчик:</b>	И.А. Макаров, старший преподаватель

## 1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-2 Способен управлять проектом на всех этапах его жизненного цикла	УК-2.2 Способен прогнозировать результат деятельности и планировать последовательность шагов для достижения данного результата. Формирует план-график реализации проекта в целом и план контроля его выполнения
	УК-2.3 Способен организовать и координировать работу участников проекта, обеспечивать работу команды необходимыми ресурсами
ОПК-2 Имеет представление об актуальных проблемах науки и техники в области информатики и вычислительной техники, способен на научном языке формулировать профессиональные задачи	ОПК-2.1 Имеет представление о современном состоянии исследований в рамках тематической области своей профессиональной деятельности
	ОПК-2.2 Способен оценивать актуальность исследований в области информатики и вычислительной техники и их практическую значимость
	ОПК-2.3 Владеет профессиональной терминологией, используемой в современной научно-технической литературе, обладает навыками устного и письменного изложения результатов научной деятельности в рамках профессиональной коммуникации
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-2.1 Знает основы научно-исследовательской деятельности в области информационных технологий, владеет знанием основ философии и методологии науки; знанием методов научных исследований и навыками их проведения
	ПК-2.2 Умеет применять полученные знания в области фундаментальных научных основ теории информации и решать стандартные задачи в собственной научно-исследовательской деятельности
	ПК-2.3 Имеет практический опыт научно-исследовательской деятельности в области информационно-коммуникационных технологий
	ПК-2.4 Владеет методами и алгоритмами решения задач цифровой обработки сигналов, использования сети Интернет, аннотирования, реферирования, библиографического поиска, опыт работы с научными источниками

## 2. Показатели оценивания компетенций

В результате изучения дисциплины «Структурный анализ и визуализация сетей» обучающийся должен:

### знать:

- знать основные понятия и терминологию, используемые в науке о сетях;
- знать основополагающие принципы сетевой структуры и эволюции.

### уметь:

- уметь разрабатывать математические модели сетевых процессов;
- уметь анализировать сетевые данные реального мира.

### владеть:

- владеть основополагающими принципами сетевой структуры и эволюции.

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

Примеры проектов, выдаваемых в течение семестра (все проекты ставят цель предсказать некоторую информацию о пользователе социальной сети):

- предсказать должность, возраст, пол и место проживания сотрудника;
- предсказать научного руководителя для научного сотрудника по базе публикаций;
- оценка области научных интересов по базе публикаций;
- по профилю автора найти среднее кол-во соавторов и индекс цитирования;
- рекомендательная система для конференций и журналов, подходящих для публикации;
- импорт данных и преобразование их в граф;
- основные характеристики графов;
- визуализация графа с использованием библиотеки matplotlib;
- степень узла и распределение степеней узлов;
- кластеризация и выделение сообществ;
- анализ географической и пространственной информации.

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

Представлены в прикрепленном файле.

##### **Критерии оценивания**

Оценка "отлично" (10 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" (9 баллов) выставляется студенту, показавшему систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений.

Оценка "отлично" (8 баллов) выставляется студенту, показавшему глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, правильное обоснование принятых решений.

хорошо

Оценка "хорошо" (7 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" (6 баллов) выставляется студенту, если он знает материал, грамотно 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 знает основной материал, грамотно излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач неточности.

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 владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации.

Оценка "удовлетворительно" (3 балла) выставляется студенту, показавшему достаточный характер знаний для дальнейшего обучения и студент может применять полученные знания по образцу в стандартной ситуации.

неудовлетворительно

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

Оценка "неудовлетворительно" (1 балл) выставляется студенту, который не знает формулировок основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач.



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

Оценка состоит из двух проектов и промежуточного контроля в конце третьего модуля. Студенты должны продемонстрировать свои знания по каждой теме лекций, касающейся как теоретических фактов, так и практических задач. Все задачи проходят через всю дисциплину и усложняются в течение курса.

Итоговая проверка проходит в виде итогового письменного экзамена. Студенты должны продемонстрировать знание теоретических фактов, но большинство задач будет оценивать их способность решать практические примеры, где студенты должны показать навыки быстрого решения базовых задач.

Формула оценивания:

Экзамен будет состоять из 10 задач, по 10 баллов за каждое, всего 100 баллов за экзамен. По завершении подготовительной части студенты представят свои заключительные проекты и защитят свои тезисы широкой аудитории. На защите проектов присутствуют приглашенные профессора.

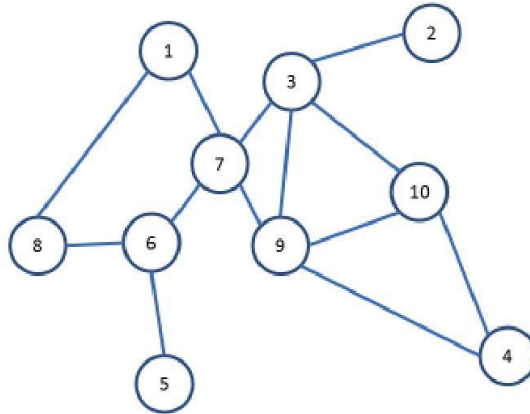
Итоговая оценка по курсу вычисляется по следующей формуле:

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

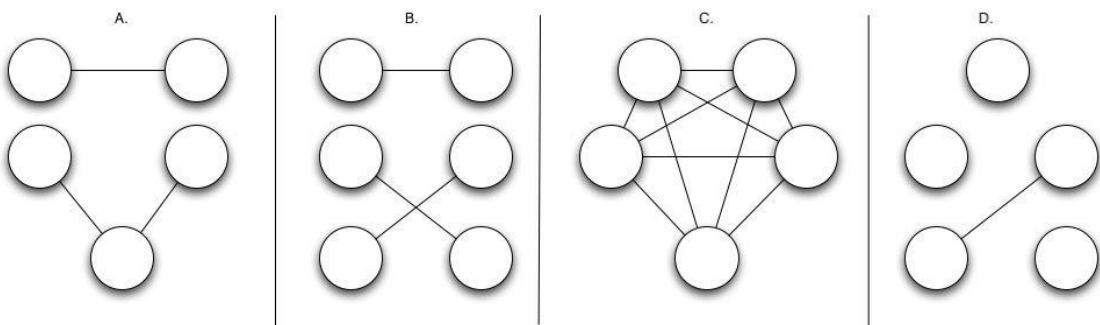
Оценки округляются в пользу экзаменатора/лектора в соответствии с регулярностью посещаемости занятий и качества второго проекта. Все оценки, имеющие дробную часть больше 0,5, округляются в большую сторону.

## Примеры вопросов к экзамену

1. Некоторая социальная сеть имеет экспоненциальное распределение степеней вершин  $P(k) = Ce^{-ak}$ , где  $C$  и  $a$  – константы,  $a$  известно. Найдите долю сетевых узлов, у которых не более  $k_0$  соседей. Используйте непрерывное приближение  $k_{\min} = 0$ ,  $k_{\max} = \infty$ .
2. Рассмотрим случайное блуждание по неориентированному графу, приведенному ниже. Вычислите вероятность попадания в узел # 9 в пределе по времени



3. Вычислите распределение степени узла в модели Барабаси-Альберта с равномерной вероятностью прикрепления  $P(k) = 1/(m_0 + T)$ .
4. Какой из графов, представленных ниже, наиболее вероятно получен с помощью Ердоша-Рени модели с параметрами  $N=5$  и  $p=0.3$ ? Почему? Какой граф не может быть сгенерирован этой моделью?



5. Для неориентированного графа-треугольника вычислите изменение модулярности при разделении одного сообщества на два.
6. Задан граф, соответствующий социальной сети, узлами которой являются люди определенного пола: 0 - мужчины, 1 - женщины. Проверьте парадокс дружбы: "В среднем ваши друзья более популярны, чем вы". Рассчитайте соотношение между средним количеством друзей и средним количеством друзей друзей друзей. Рассчитайте соотношение отдельно для мужчин и женщин. Пожалуйста, обратите внимание, что друзьями мужчин могут быть как мужчины, так и женщины, а также подруги женщин.

7. Задан граф, узлами которой являются люди определенного возраста. Найдите топ-5 авторитетов и топ-5 хабов с помощью алгоритма HITS, если границы направлены от более молодых людей к более старшим.
8. Приведите пример графа с произвольным числом вершин, диаметр которого в десять раз превышает среднюю длину кратчайшего пути.
9. Для заданного графа  $G = \{\{1,2,3\}, \{(1,2), (1,3), (2,3)\}\}$  вычислите метрику HITS для узлов  $\{1,2,3\}$ .
10. Дан граф, узлами которой являются люди определенного возраста. Найдите топ-5 пользователей по PageRank, где матрица переходов пропорциональна возрасту узла назначения. Вероятность перехода равна 0,85, вероятность телепортации равна 0,15.
11. Дан граф. Найдите оптимальное количество сообществ по максимальному разрыву в собственных значениях графа Лапласиана. Итеративно разделите граф на основе собственных векторов Лапласиана графа и укажите количество узлов в самом большом сообществе.
12. Дан граф, узлами которой являются люди с определенным регионом. Постройте векторные представления узлов 16d, используя усеченный SVD графа Лапласиана. Разделите узлы по регионам и рассчитайте оценку силуэта в sklearn.