

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ**

федеральное государственное автономное образовательное учреждение
высшего образования

**«Московский физико-технический институт
(национальный исследовательский университет)»
(МФТИ)**

УТВЕРЖДАЮ

Ректор МФТИ

д-р физ.-мат. наук, профессор

Д.В. Ливанов

«31»

мая

2023 г.



**Программа
дополнительного образования
«Веб-графы и методы работы с ними»**

Москва 2023

1. Общая характеристика программы

1.1. Целью реализации дополнительной общеобразовательной программы «Веб-графы и методы работы с ними» является изучение теории графов и их применения к веб-разработке, а также разработка навыков построения, анализа и оптимизации веб-графов.

1.2. Программа предназначена для специалистов в области информационно-коммуникационных технологий, имеющих высшее образование; студентов вузов.

1.3. Нормативный срок освоения программы – 72 академических часов.

1.4. Форма обучения – очно-заочная с применением дистанционных образовательных технологий.

Программа может быть реализована в сетевой форме.

1.5. Режим обучения: 8 недель (9 часов в неделю).

2. Планируемые результаты обучения

В результате освоения программы слушатель должен:
знать:

- основные понятия теории графов и их применение в веб-разработке,
- основные методы построения веб-графов и их использование для анализа и оптимизации веб-сайтов,

уметь:

- создать веб-граф для заданного сайта, проанализировать его структуру,
- оптимизировать веб-сайт на основе веб-графа,
- разрабатывать алгоритмы обхода и поиска в веб-графах.

3. Структура программы

Программа предусматривает изучение следующих тем (модулей):

1. Свойства веб-графов и сложных сетей
2. Модели Боллобаша-Риордана и Бакли-Остгуса
3. Количество ребер между вершинами заданных степеней.
Ассортативность. Линковые кольца. Соответствие модели Бакли-Остгуса реальному хост-графу
4. Кластерные коэффициенты. Число копий фиксированного графа
5. Виды центральностей и PageRank
6. Обзор перспективных моделей. Модель свежести, модель копирования.
РА-класс.

Структура программы представлена в таблице 1.

Таблица 1

№	Тема (модуль)	Кол-во часов	В том числе	
			Аудит. занятия	Самост. работа
0	Дополнительный материал по Python	5,2	1,3	3,9
1	Свойства веб-графов и сложных сетей	9,8	1,8	8
2	Модели Боллобаша-Риордана и Бакли-Остгуса	12,2	2,7	9,5
3	Количество ребер между вершинами заданных степеней. Ассортативность. Линковые кольца. Соответствие модели Бакли-Остгуса реальному хост-графу	9,7	1,7	8
4	Кластерные коэффициенты. Число копий фиксированного графа	9,9	1,9	8
5	Виды центральностей и PageRank	11,3	2,3	9
6	Обзор перспективных моделей. Модель свежести, модель копирования. РА-класс	11,9	2,5	9,4
	Экзамен	2	1	1
Итого		72	15,2	56,8

4. Содержание программы

4.1. Учебно-тематический план программы

Таблица 2

Тема (модуль)	Тема урока	Кол-во часов	
		Аудит. занятия	Самост. работа
0. Дополнительный материал по Python	Введение в язык Python: типы данных	0,4	1,2
	Введение в язык Python: словари, массивы и множества	0,1	0,3
	Введение в язык Python: условные	0,2	0,6

	операторы и циклы		
	Введение в язык Python: функции и генераторы	0,2	0,6
	Введение в библиотеку networkX	0,3	0,9
	Библиотека networkX: связность графов	0,1	0,3
1. Свойства веб-графов и сложных сетей	Сложные сети и их свойства	0,4	0,5
	Гигантская компонента	0,1	0,1
	Диаметр сложных сетей	0,2	0,2
	Устойчивость и уязвимость к атакам на хабы	0,3	0,4
	Распределение степеней вершин	0,2	0,2
	Решение задачи "распределение степеней вершин"	0,2	0,2
	Решение задачи "устойчивость и уязвимость"	0,2	0,2
	Исследование фрагмента хост-графа	0,2	0,2
	Задание на программирование 1		4
	Тест 1		2
2. Модели Боллобаша-Риордана и Бакли-Остгуса	"Модель" Барабаша-Альберт	0,2	0,3
	Критика модели Барабаша-Альберт	0,1	0,1
	Определение модели Боллобаша-Риордана	0,3	0,4
	Статическое определение модели	0,3	0,4
	Свойства случайных графов в модели Боллобаша-Риордана	0,3	0,4
	Распределение степеней вершин	0,2	0,2
	Модель Бакли-Остгуса	0,4	0,5
	Генерация графа в модели Боллобаша-Риордана	0,2	0,3
	Распределение вершин в модели Боллобаша-Риордана	0,1	0,1
	Устойчивость гигантской компоненты	0,2	0,3
	Генерация графа в модели Бакли-Остгуса	0,2	0,3
	Наименьшая степень вершины	0,2	0,2
	Задание на программирование 2		4
	Тест 2		2
3. Количество ребер между вершинами заданных степеней. Ассортативность. Линковые кольца.	Линковые кольца	0,2	0,3
	Количество ребер между вершинами заданных степеней	0,2	0,3
	Количество ребер между вершинами заданных степеней в модели Бакли-Остгуса	0,2	0,2
	Соответствие модели Бакли-Остгуса реальному хост-графу	0,3	0,4
	Классификация ссылочного спама	0,1	0,1
	Корреляция степеней вершин и	0,2	0,2

Соответствие модели Бакли-Остгуса реальному хост-графу	ассортативность		
	Ассортативность графа в модели Бакли-Остгуса	0,1	С,1
	Количество ребер между вершинами заданных степеней в модели Бакли-Остгуса	0,2	С,2
	Аппроксимация полученных результатов	0,2	С,2
	Задание на программирование 3		4
	Тест 3		2
4. Кластерные коэффициенты. Число копий фиксированного графа	Глобальный и средний локальный кластерные коэффициенты	0,2	С,2
	Кластерный коэффициент: пример Боллобаши	0,2	С,2
	Глобальный кластерный коэффициент	0,2	С,2
	Кластерный коэффициент в модели Боллобаши-Риордана	0,3	0,4
	Кластерный коэффициент в модели Бакли-Остгуса	0,1	0,1
	Цепи и клики в модели Бакли-Остгуса	0,2	0,2
	Циклы и полные двудольные графы в модели Бакли-Остгуса	0,2	0,2
	Кластерные коэффициенты	0,3	0,3
	Треугольники в модели Боллобаши-Риордана	0,1	0,1
	Соотношение глобального и локального кластерных коэффициентов	0,1	0,1
	Задание на программирование 4		4
	Тест 4		2
5. Виды центральностей и PageRank	Центральность betweenness	0,2	0,3
	Центральность closeness	0,1	0,1
	Центральность eigenvector	0,1	0,1
	Пейджеранк	0,3	0,5
	Теорема Авраченко в конкретизации модели предпочтительного присоединения	0,3	0,5
	Weighted Yandex PageRank	0,3	0,5
	PageRank	0,1	0,1
	Центральности	0,2	0,3
	Pagerank и виды центральностей	0,2	0,3
	Pagerank в модели Авраченко	0,3	0,5
	Накрутка PageRank'a	0,2	0,3
	Задание на программирование 5		4
	Тест 5		2
6. Обзор перспективных	РА-класс моделей	0,3	0,4
	Степенной закон и кластерные коэффициенты	0,2	0,3

моделей. Модель свежести, модель копирования. РА-класс	Модели свежести	0,3	0,4
	Свойства модели свежести	0,1	0,1
	Модель Боллобаша-Боргса-Риордана-Чайес	0,2	0,3
	Модель копирования	0,3	0,4
	Модель Боллобаша-Боргса-Риордана-Чайес	0,3	0,4
	Модель копирования	0,2	0,3
	РА-условие в модели Холма-Кима	0,3	0,4
	РА-условие в модели Мори	0,2	0,3
	Ограниченный локальный кластерный коэффициент	0,1	0,1
	Задание на программирование 6		4
	Тест 6		2
Экзамен		1	1

4.2. Учебная программа по модулям

Таблица 3

№ п/п	Наименование темы (модуля)	Содержание обучения, наименование и тематика практических занятий (семинаров), самостоятельной работы, используемых образовательных технологий
0	Дополнительный материал по Python	Практические занятия Введение в язык Python: типы данных Введение в язык Python: словари, массивы и множества Введение в язык Python: условные операторы и циклы Введение в язык Python: функции и генераторы Введение в библиотеку networkX Библиотека networkX: связность графов Самостоятельная работа Самостоятельное выполнение заданий по теме лекции, изучение дополнительных материалов
1	Свойства веб-графов и сложных сетей	Лекции Сложные сети и их свойства Гигантская компонента Диаметр сложных сетей Устойчивость и уязвимость к атакам на хабы Практические занятия Распределение степеней вершин Решение задачи "распределение степеней вершин" Решение задачи "устойчивость и уязвимость" Исследование фрагмента хост-графа Самостоятельная работа Самостоятельное выполнение заданий на программирование по теме лекции, изучение дополнительных материалов, тестирование.
2	Модели Боллобаша-	Лекции

	Риордана и Бакли-Остгуса	"Модель" Барабаши-Альберт Критика модели Барабаши-Альберт Определение модели Боллобаша-Риордана Статическое определение модели Свойства случайных графов в модели Боллобаша-Риордана Практические занятия Распределение степеней вершин Модель Бакли-Остгуса Генерация графа в модели Боллобаша-Риордана Распределение вершин в модели Боллобаша-Риордана Устойчивость гигантской компоненты Генерация графа в модели Бакли-Остгуса Наименьшая степень вершины Самостоятельная работа Самостоятельное выполнение заданий на программирование по теме лекции, изучение дополнительных материалов, тестирование.
3	Количество ребер между вершинами заданных степеней. Ассортативность. Линковые кольца. Соответствие модели Бакли-Остгуса реальному хост-графу	Лекции Линковые кольца Количество ребер между вершинами заданных степеней Количество ребер между вершинами заданных степеней в модели Бакли-Остгуса Соответствие модели Бакли-Остгуса реальному хост-графу Классификация ссылочного спама Корреляция степеней вершин и ассортативность Практические занятия Ассортативность графа в модели Бакли-Остгуса Количество ребер между вершинами заданных степеней в модели Бакли-Остгуса Аппроксимация полученных результатов Самостоятельная работа Самостоятельное выполнение заданий на программирование по теме лекции, изучение дополнительных материалов, тестирование.
4	Кластерные коэффициенты. Число копий фиксированного графа	Лекции Глобальный и средний локальный кластерные коэффициенты Кластерный коэффициент: пример Боллобаша Глобальный кластерный коэффициент Кластерный коэффициент в модели Боллобаша-Риордана Кластерный коэффициент в модели Бакли-Остгуса Цепи и клики в модели Бакли-Остгуса Циклы и полные двудольные графы в модели Бакли-Остгуса Практические занятия Кластерные коэффициенты

		Треугольники в модели Боллобаша-Риордана Соотношение глобального и локального кластерных коэффициентов Самостоятельная работа Самостоятельное выполнение заданий на программирование по теме лекции, изучение дополнительных материалов, тестирование.
5	Виды центральностей и PageRank	Лекции Центральность betweenness Центральность closeness Центральность eigenvector Пейджеранк Теорема Авраченко в конкретизации модели предпочтительного присоединения Weighted Yandex PageRank Практические занятия PageRank Центральности Pagerank и виды центральностей Pagerank в модели Авраченко Накрутка PageRank'a Самостоятельная работа Самостоятельное выполнение заданий на программирование по теме лекции, изучение дополнительных материалов, тестирование.
6	Обзор перспективных моделей. Модель свежести, модель копирования. РА-класс	Лекции РА-класс моделей Степенной закон и кластерные коэффициенты Модели свежести Свойства модели свежести Модель Боллобаша-Боргса-Риордана-Чайес Модель копирования Практические занятия Модель Боллобаша-Боргса-Риордана-Чайес Модель копирования РА-условие в модели Холма-Кима РА-условие в модели Мори Ограниченный локальный кластерный коэффициент Самостоятельная работа Самостоятельное выполнение заданий на программирование по теме лекции, изучение дополнительных материалов, тестирование.
	Экзамен	Практическое занятие: итоговое тестирование по материалам лекций Самостоятельная работа: подготовка к итоговому

Примеры заданий для организации самостоятельной работы слушателей

Задание на программирование

Задача 1.

Для каждого $d \in \mathbb{N}$ пусть $g(d)$ – доля достижимых друг из друга пар вершин заданного графа, расстояние между которыми не более, чем d . Построим следующий график: отметим точки $\{(d, g(d))\}$ для всех d , и каждую пару точек $(d, g(d))$ и $(d + 1, g(d + 1))$ соединим прямой, их соединяющей. В результате получится график некоторой кусочно-линейной функции. Эффективный диаметр определяется как значение $\tilde{d}$ при котором полученная функция равна 0.9.

В файле [1.graph.txt](#) находится фрагмент веб-графа, заданный списком рёбер. В каждой строке через символ табуляции задана пара вершин, соединённых ребром.

Найдите диаметр и эффективный диаметр самой большой компоненты связности этого графа.

Ответом на задачу являются два числа - диаметр и эффективный диаметр самой большой компоненты связности графа. Округлите второе число до сотых.

Для вычисления расстояний между всеми парами вершин в графе воспользуйтесь функцией `networkx.shortest_path_length`. Она возвращает словарь, ключи которого – вершины графа, а значения – словари расстояний. То есть, если результат вызова – словарь `dist`, то `dist[u][v]` – расстояние от вершины u до вершины v в графе.

Задача 2.

Сгенерируйте графы в такой модели при различных значениях t до $t = 3000$ (с каким-нибудь шагом) и эмпирически установите, чему в пределе равна доля вершин степени 2.

Ответом на задачу является единственное число, округлите его до десятых.

Рассмотрим следующую модель генерации неориентированных графов. Граф $G^{(0)}$ состоит из одного ребра. Для всех $t, t \geq 1$ граф $G^{(t)}$ получается из графа $G^{(t-1)}$ добавлением новой вершины, соединённой с обоими концами случайного ребра, выбранного равновероятно из множества всех рёбер

графа $G^{(t-1)}$. Таким образом, граф $G^{(1)}$ – это треугольник, граф $G^{(2)}$ состоит из 4 вершин и 5 ребер и т.д.

Задача 3.

В файле 3.graph.txt задан ориентированный граф – в каждой строке через символ табуляции записана пара вершин $u \langle tab \rangle v$, соответствующая ребру из u в v в графе.

Постройте график ассортативности $d_{nn}(d)$ данного графа. С помощью вызова `curve_fit` без специальных параметров аппроксимируйте эту функцию функцией вида $c \cdot d^\delta$. Необходимо использовать все пары $(d, d_{nn}(d))$. Постройте график и полученную аппроксимацию в log-log координатах. Является ли граф ассортативным?

Обратите внимание, что в графе есть петли и кратные ребра.

Ответом на задачу является единственное число – оптимальное значение параметра округленное до сотых.

Задача 4.

Найдите $\#(K_{3,3}, G)$ для неориентированного графа G , заданного в файле 4.graph.txt списком ребер.

Дополнительно сгенерируйте граф $H_{a,m}^n$ в модели Бакли-Остгуса при $n = 2000, m = 3, a = \frac{1}{3}$ и найдите $\#(K_{3,3}, H_{a,m}^n)$ для полученного графа. Сравните с величиной $\#(K_{3,3}, G_2)$ для графа G_2 в модели Эрдеша-Реньи на том же количестве вершин и с близким количеством ребер.

Ответом на задачу является единственное целое число – число подграфов $K_{3,3}$ в графе из файла.

Напишите функцию поиска в заданном неориентированном графе подграфов $K_{3,3}$.

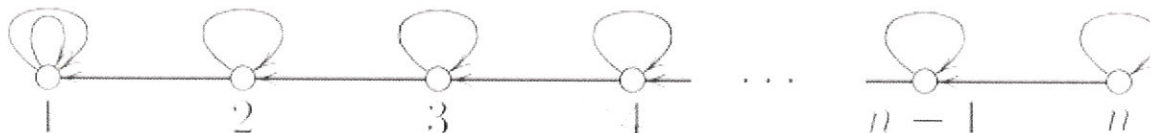
Тестовые вопросы

Вопрос 1. Чему равен диаметр графа на рисунке?

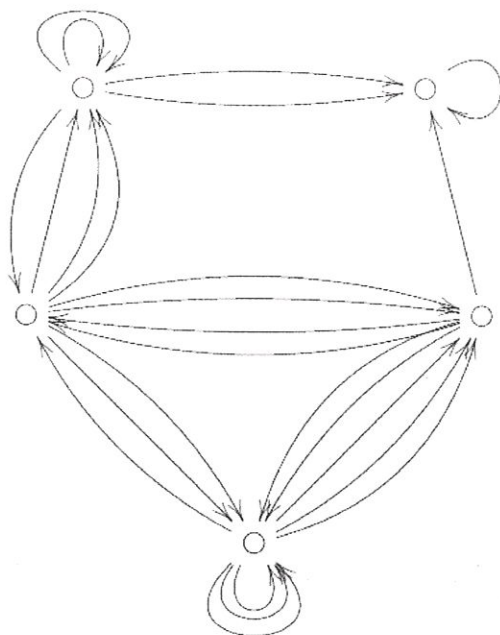


Вопрос 2. Сколько ребер в графе $H_{1/3.12}^{3000}$ в модели Бакли-Остгуса ($a = \frac{1}{3}, m = 12, n = 3000$)?

Вопрос 3. Найдите вероятность того, что случайный граф G_2^n Боллобаша-Риордана совпадет с графом на рисунке:



Вопрос 4. Найдите значение величины $X(10, 10) + X(10, 11) + X(11, 14)$ для графа на рисунке:



4.3. Список рекомендуемой литературы

1. Ахо А., Ульман Д. Теория синтаксического анализа, перевода и компиляции. М.: Мир, 2008.
2. Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К. Алгоритмы: построение и анализ. М.: МТ Пресс, 2006.
3. Безрукавный А. Веб-аналитика. Как анализировать посетителей сайта. М.: ДМК Пресс, 2015.
4. О'Рейли Т. Мастеринг алгоритмов с помощью Python. М.: ДМК Пресс, 2017.

5. Брин С., Пейдж Л. Анатомия крупномасштабного гипертекстового поискового движка, <https://efim360.ru/anatomiya-krupnomasshtabnogo-gipertekstovogo-poiskovogo-dvizhka-sergej-brin-lourens-pejdzh/>

5. Материально-технические условия реализации программы

Таблица 4

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
Система дистанционного обучения	Аудиторные занятия	Слушателю необходим компьютер, наличие доступа в сеть интернет.
Система дистанционного обучения	Самостоятельная работа	Слушателю необходим компьютер, наличие доступа в сеть интернет.
Система дистанционного обучения	Экзамен	Слушателю необходим компьютер, наличие доступа в сеть интернет.

6. Оценка качества освоения программ

Оценка качества освоения программы осуществляется в процессе промежуточной аттестации.

Формы и методы промежуточного контроля представлены в таблице 5.

Таблица 5

Наименование модулей	Основные показатели оценки	Формы и методы контроля и оценки
Дополнительный материал по Python	Установленное количество выполненных заданий	Самостоятельное выполнение заданий
Свойства веб-графов и сложных сетей	Установленное количество выполненных заданий	Самостоятельное выполнение заданий по программированию, тестирование
Модели Боллобаша-Риордана и Бакли-Остгуса	Установленное количество выполненных заданий	Самостоятельное выполнение заданий по программированию, тестирование
Количество ребер между вершинами заданных степеней. Ассортативность. Линковые кольца. Соответствие модели Бакли-	Установленное количество выполненных заданий	Самостоятельное выполнение заданий по программированию, тестирование

Остгуса реальному хост-графу		
Кластерные коэффициенты. Число копий фиксированного графа	Установленное количество выполненных заданий	Самостоятельное выполнение заданий по программированию, тестирование
Виды центральностей и PageRank	Установленное количество выполненных заданий	Самостоятельное выполнение заданий по программированию, тестирование
Обзор перспективных моделей. Модель свежести, модель копирования. РА-класс	Установленное количество выполненных заданий	Самостоятельное выполнение заданий по программированию, тестирование
Экзамен	Обобщение знаний по темам курса. Применение полученных знаний.	Тестирование

7. Составители программы

Зубцова Жанна Исхаковна, к.ф.-м.н., ведущий специалист Отдела сопровождения образовательных программ (Центр «Пуск»).

Согласовано
Ведущий специалист ОСОП

«19» 05 Ж. И. Зубцова
2023 г.

Согласовано
Руководитель проектов (Центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск")

«19» 05 М. В. Выголова
2023 г.