



Санкт-Петербургский  
государственный университет

Доклад в Санкт-Петербургский  
институт информатики  
и автоматизации  
Российской академии наук  
(СПИИРАН)

19 апреля 2024 г.

Герштейн А.М. «Программные инструменты для построения безопасных маршрутов транспорта»,  
диссертация на соискание ученой степени кандидата технических наук.  
Научный руководитель зав. кафедрой системного программирования СПбГУ,  
д.ф.м.н., профессор Терехов А.Н.



Автор выражает благодарность Терехову А.Н.  
за научное руководство, поддержку  
и помощь при работе над диссертацией.

Кафедре Математической Механики СПбГУ  
за обучения в Аспирантуре по Информатике.

Искандерову, Ю.М. за приглашение выступить  
здесь с сегодняшним докладом.



# 1. Практическая значимость

Согласно данным WHO \*

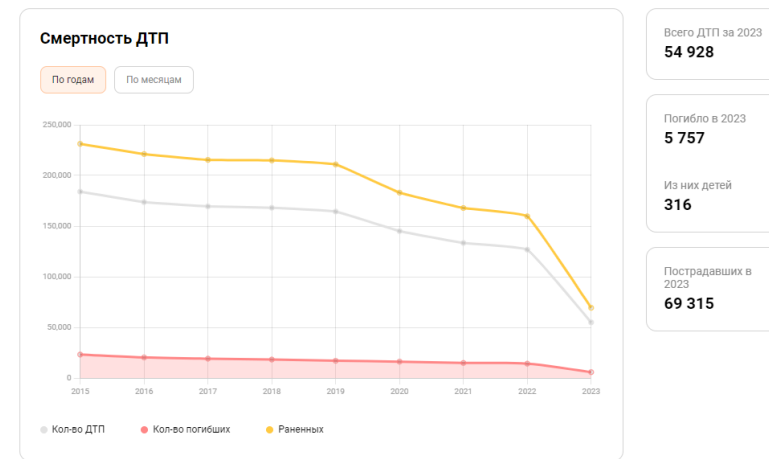
(World Health Organization) за год около

**1.3 миллиона человек** во всем мире **погибает**

в дорожных авариях и примерно

**50 миллионов** получают **травмы** различной степени тяжести.

Автором создан пакет программ на языке Python, позволяющий модифицировать маршрут транспортного средства таким образом, что он будет в среднем безопасней оригинального маршрута.



## 2. Представление дорожной сети

При прокладке маршрута используется представление **дорожной сети** в виде **направленного графа**, где **сегменты** дороги заменяются **ребрами** графа (ребра могут быть **одно- и двунаправленными**) а **вершины** графа обозначают **переход от одного сегмента к другому**. На Рис. 1 показан пример дорожного графа с ребрами и вершинами, помеченными красными точками.



Рис. 1. Фрагмент дорожной сети Ньютона (Массачусетс)



### 3. Более безопасный маршрут

Обычные алгоритмы маршрутизации (**Дейкстра**) не учитывают безопасность маршрута, стремятся построить как можно более быстрый маршрут. Каким же образом сделать маршрут более безопасным? Общий подход, реализованный в диссертации, состоит в том, чтобы выделить на соответствующем графе участки повышенной опасности и затем модифицировать дорожный граф таким образом, чтобы алгоритм маршрутизации обходил эти участки.

### 4. Первый подход к выделению УПО — кластеризация ДТП

Для нахождения кластеров используется алгоритм **DBSCAN** — эвристический алгоритм, находящий кластеры одинаковой минимальной плотности (буквы DB в названии алгоритма означают **Dense-Based**, т.е. ориентированный на плотность) и произвольной формы.

Кластеризация проводилась отдельно для каждого года из 6 (2013-2018) с параметрами:

- минимальное число элементов в кластере=3,
- максимальное расстояние между элементами в кластере = 10м

## 4. Первый подход к выделению УПО – продолжение

Выделить статистически значимые кластеры.

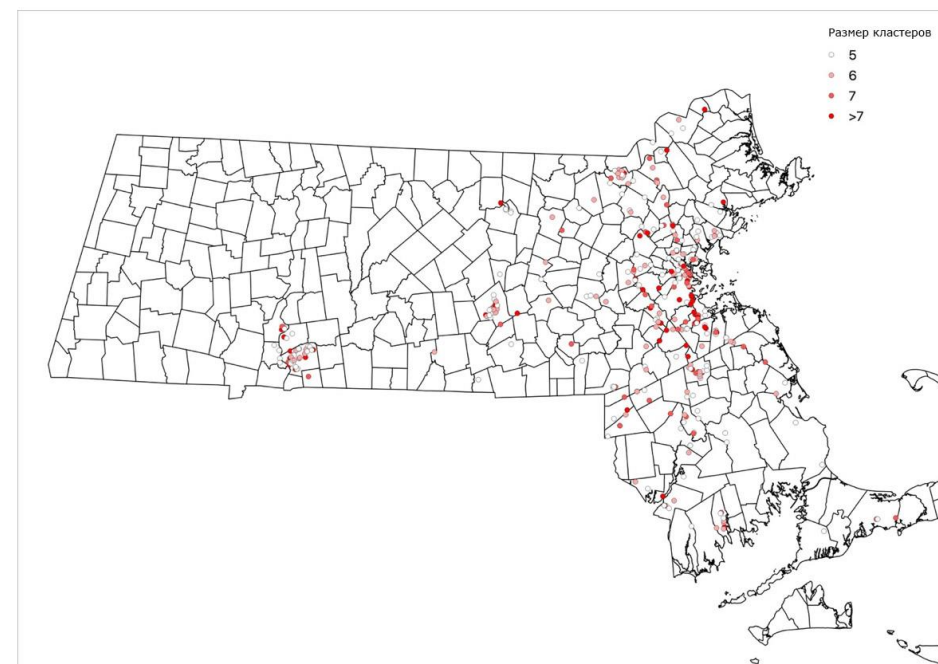
- Выделить статистически значимые кластеры с помощью *статистических испытаний*.
- Для статистических испытаний *использовать столько же точек, сколько зарегистрировано ДТП, но равномерно распределенных по дорожной сети*.
- Произвести несколько сот статистических испытаний, далее произвести кластеризацию для каждого набора точек и *собрать статистику размеров кластеров*.
- На основании статистики, полученной в результате статистических испытаний, *оставить только те реальные кластеры, чьи размеры маловероятны в случае равномерного распределения ДТП по дорожной сети*

Таблица 1. Результаты моделирования методом Монте-Карло для Массачусетса, 2013 год (1502 испытания, 23964 точки в каждом)

Размер	Количество кластеров одинакового или большего размера	P
3	1502	1
4	160	0.1
5	6	0.004

Таким образом, на уровне 0,05 следует рассматривать кластеры размером  $\geq 5$  как статистически значимые.

РИС. 2. Массачусетс 2013, значимые кластеры серьезных ДТП (DBSCAN, eps=10)



## 4-1. Кластеризация — повторяемость

Для каждого кластера ДТП, случившихся в одном году найдем эквивалентный (с расстоянием между двумя центрами кластеров  $<10m$ ) кластер ДТП, случившихся в другом году. Результаты представлены в Таблице 1.

- Число 0.36 на пересечении столбца 2017 и строки 2018 означает, что отношение (количество кластеров в 2017 году, найденных в 2018 году/общее количество кластеров в 2017 году) равно 0,36

Таблица 2. Относительная повторяемость кластеров из года в год.

Годы	2017	2016	2015	2014	2013	Кластеры
2018	0.36	0.35	0.35	0.33	0.36	378
2017		0.35	0.35	0.34	0.32	369
2016			0.39	0.33	0.3	374
2015				0.3	0.33	334
2014					0.36	349
2013						354

## 5. Оценка эффективности маршрутизации

После того как проложен маршрут по модифицированному графу возникает следующая проблема: *как оценить эффективность новой маршрутизации, насколько она делает движение безопасней?*

- Для этого, во-первых, нужно каким-то образом *создать множество маршрутов, покрывающих весь дорожный граф*, и,
- во-вторых, *использовать показатель средней эффективности новой маршрутизации*.

Предложенное решение:

- Создать множество маршрутов, покрывающих весь дорожный граф, *для чего использовать квадратную сетку, наложенную на дорожный граф*.
- Маршруты из каждого узла решетки в каждый, исключая одинаковые начало и конец. *Перебирая все узлы решетки, за исключением одинаковых, получим множество маршрутов типа «вершина, ближайшая к начальному узлу решетки» - «вершина, ближайшая к конечному узлу решетки»*.
- *Для каждого такого маршрута подсчитать относительный риск ДТП (ОРДТП), равный отношению числа ДТП вдоль модифицированного маршрута к числу ДТП вдоль оригинального маршрута, построенного по не измененному графу*.
- *Если среднее значение ОРДТП значительно меньше единицы, будем считать, что модифицированный маршрут безопасней оригинального*.

## 6. МОДЕЛИРОВАНИЕ МНОЖЕСТВА ПУТЕЙ ПУТЕМ НАЛОЖЕНИЯ КВАДРАТНОЙ СЕТКИ НА ДОРОЖНУЮ СЕТЬ

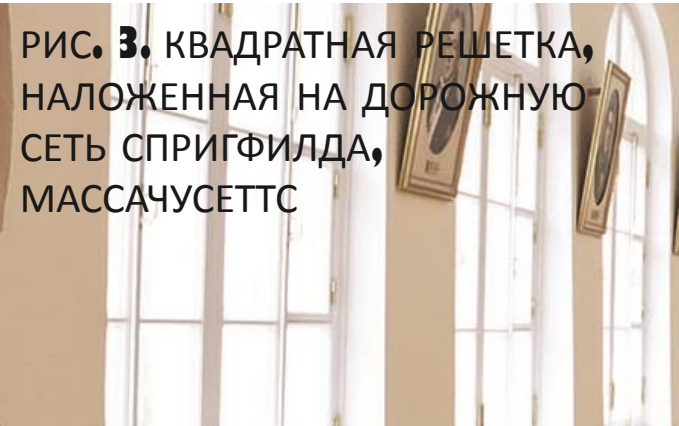


РИС. 3. КВАДРАТНАЯ РЕШЕТКА, НАЛОЖЕННАЯ НА ДОРОЖНУЮ СЕТЬ СПРИГФИЛДА, МАССАЧУСЕТТС

- На Рис. 3 показана квадратная решетка, наложенная на карту Спрингфилда (Массачусетс). На карте поместилось 95 узлов сетки с расстоянием около 830 м. между соседними узлами, что дает  $95 \cdot 94 = 8930$  различных маршрутов.
- При этом *маршрут A → B не эквивалентен маршруту B → A, поскольку существуют однонаправленные ребра дорожного графа.*
- Теперь для каждой пары вершин вычислим *относительный риск ДТП (ОРДТП)* как *отношение числа ДТП вдоль измененного маршрута к числу ДТП вдоль оригинального маршрута* (т.е. полученного на первоначальном, не модифицированном графе).
- Для Москвы и СПб - данные за 3 года, для США - данные за 6 лет.

Если среднее всех таких отношений значительно меньше единицы, то можно сказать, что маршруты, вычисленные на модифицированном графе, в среднем **безопасней**, чем маршруты, полученные на оригинально графе. То есть верна статистическая гипотеза, что средний ОРДТП  $< 1$ .

Имеет смысл усреднять ОРДТП лишь для определенных длин не модифицированного маршрута, что сделает картину более детальной.



## 7. Зависимость относительного риска ДТП от длины оригинального (без учета кластеров) маршрута для города Спрингфилд (Массачусетс)



Чтобы создать модифицированный граф, нужно было рассмотреть каждый кластер и определить ребра, ведущие к нему.

У каждого такого ребра менялся атрибут длины, который полагался равным очень большому ( $10^{13}$ м), практически бесконечному числу, на много порядков превышающему максимальную длину маршрута в Спрингфилде ( $1.4 \times 10^5$ м).

**По сути, это означает запрет движения через выделенный на графе кластер ДТП.**

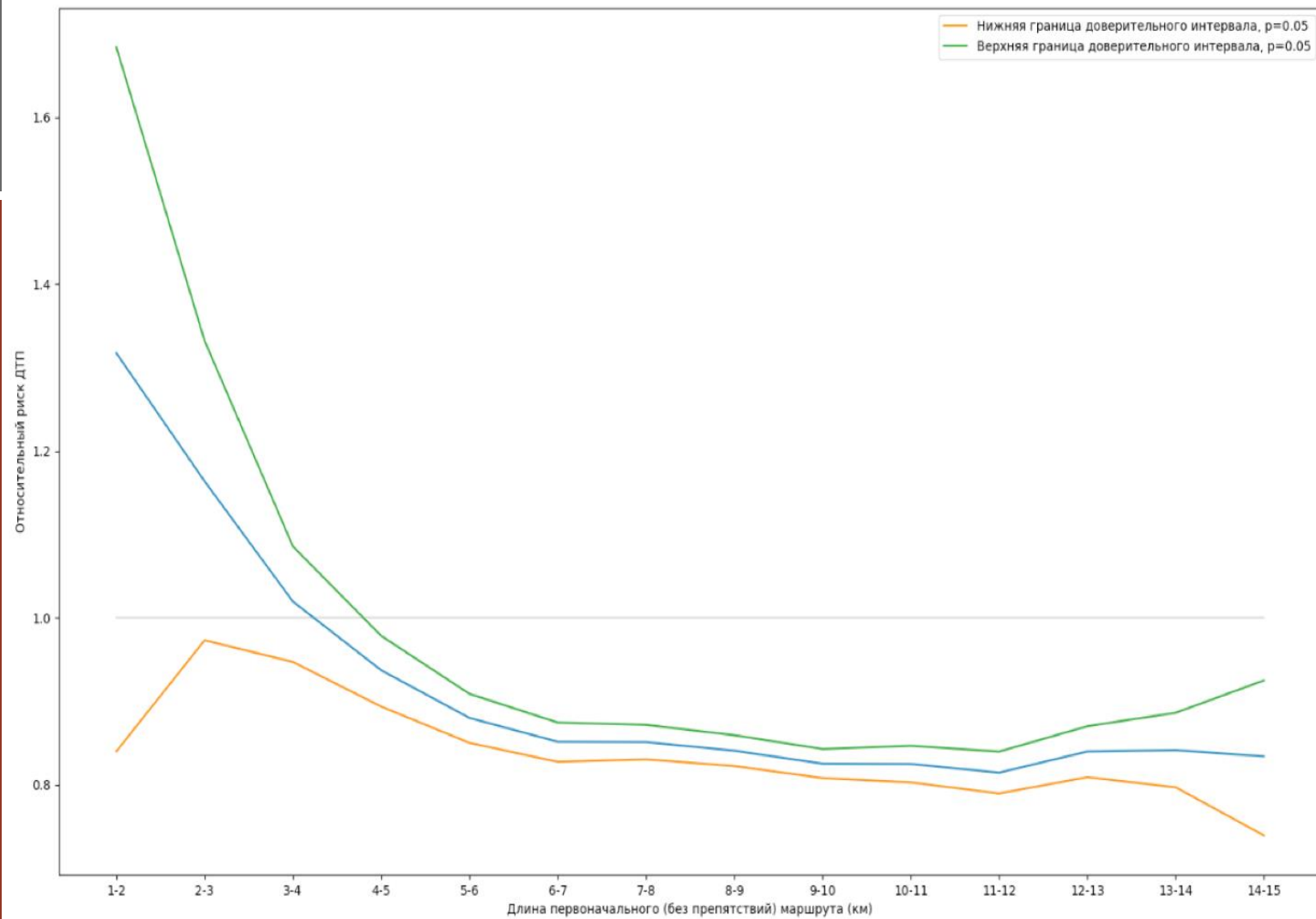
На Рис. 4: зависимость средних отношений длин маршрутов (длина при обходе упо/ первоначальная длина) для различной длины оригинального, проложенного без учета УПО, маршрута

- На Рис. 4 показана **зависимость относительного риска ДТП от длины оригинального** (без учета кластеров) маршрута для города Спрингфилд (Массачусетс).
- В качестве опасных участков, которые необходимо обойти, здесь использовались метакластеры, получаемые в два этапа:
  - сначала определялись статистически достоверные кластеры для каждого года (использовались данные за 6 лет: 2013-2018)
  - Затем центры кластеров<sup>1</sup> полученных на первом этапе, подвергались вторичной кластеризации и отбору, который оставлял только кластеры, в которых присутствуют не менее чем три различных года. Всего таких кластеров нашлось 31.

<sup>1</sup> Центр определялся как среднее арифметическое координат X и Y



**РИС. 4. ЗАВИСИМОСТЬ СРЕДНЕГО ОТНОСИТЕЛЬНОГО РИСКА ДТП (ЧИСЛО ДТП ПРИ ОБХОДЕ УПО/ ЧИСЛО ДТП ВДОЛЬ ОРИГИНАЛЬНОГО МАРШРУТА) ДЛЯ РАЗЛИЧНОЙ ДЛИНЫ ОРИГИНАЛЬНОГО, ПРОЛОЖЕННОГО БЕЗ УЧЕТА УПО, МАРШРУТА, СПРИНГФИЛД, МАССАЧУСЕТС, СТАТИСТИКА ТОЛЬКО ПО ИЗМЕНЕННЫМ МАРШРУТАМ.**



Короткие (менее 4 км) маршруты оказываются в среднем менее безопасными, чем оригинальные (средний относительный риск  $> 1$ ).

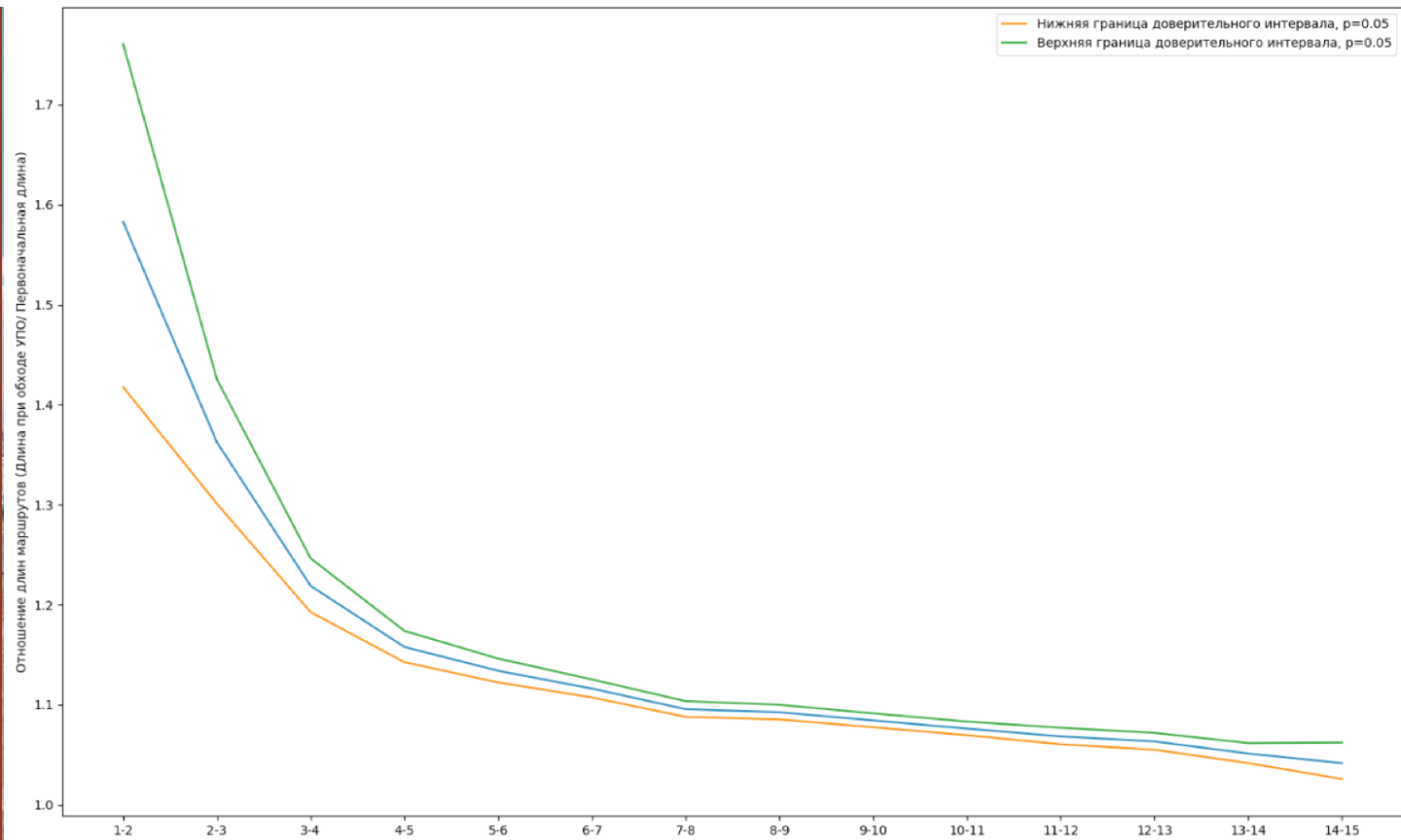
Это, очевидно, связано с тем, что на коротких маршрутах обходной путь оказывается очень длинным, как это уже было показано на Рис. 5.

С увеличением длины оригинального маршрута средний относительный риск уменьшается и стремится к значению 0.82.

Заметим, что при вычислении средних учитывались только измененные маршруты (всего их оказалось 4831 из 8930 возможных, то есть, примерно 54%).



РИС. 5. ЗАВИСИМОСТЬ СРЕДНИХ ОТНОШЕНИЙ ДЛИН МАРШРУТОВ (ДЛИНА ПРИ ОБХОДЕ УПО/ ПЕРВОНАЧАЛЬНАЯ ДЛИНА) ДЛЯ РАЗЛИЧНОЙ ДЛИНЫ ОРИГИНАЛЬНОГО, ПРОЛОЖЕННОГО БЕЗ УЧЕТА УПО, МАРШРУТА



Видно, что для коротких (меньше 4 км) маршрутов проигрыш в длине маршрута весьма существенен и достигает значения 1.58.

При увеличении длины маршрута проигрыш существенно уменьшается и равен 10% - 4%, в зависимости от длины первоначального маршрута.



## Второй подход к выделению УПО:

### 8. Дорожные сегменты как УПО

Обход кластеров, как это следует из уже сказанного, сводится к обходу выделенных дорожных сегментов. Откуда возникает вопрос: *а нельзя ли выделить «опасные» дорожные сегменты без всяких кластеров?*

**Предложенное решение:**

*Выделим те ребра* дорожного графа (= сегменты дороги), *где число ДТП превышает некий порог, определяемый статистическими испытаниями*, пометим эти ребра как опасные, т.е. *придадим им дополнительную длину (штраф)*, которая будет учитываться алгоритмом маршрутизации при прокладке маршрута из начальной точки в конечную.

- Заметим, что такой подход представляется более естественным, поскольку модифицируется непосредственно дорожный граф без привлечения дополнительных объектов (кластеров ДТП).
- *Использование кластеров связано также с объемной ручной работой: необходимо рассмотреть каждый кластер и пометить ребра дорожного графа, ведущие к нему.* Автоматизация этой задачи по меньшей мере весьма сложна.



## РИС. 6. ДОРОЖНАЯ СЕТЬ МОСКВЫ С РЕБРАМИ-ПРЕПЯТСТВИЯМИ

На каждое выделенное ребро-препятствие накладывался «штраф» — к атрибуту длины ребра прибавлялось число, что служило сигналом для алгоритма маршрутизации по возможности изменить маршрут.



## 8. Дорожные сегменты как УПО — на примере г. Москвы

На Рис. 6 показана дорожная сеть Москвы с выделенными на ней дорожными сегментами (ребрами) — препятствиями.

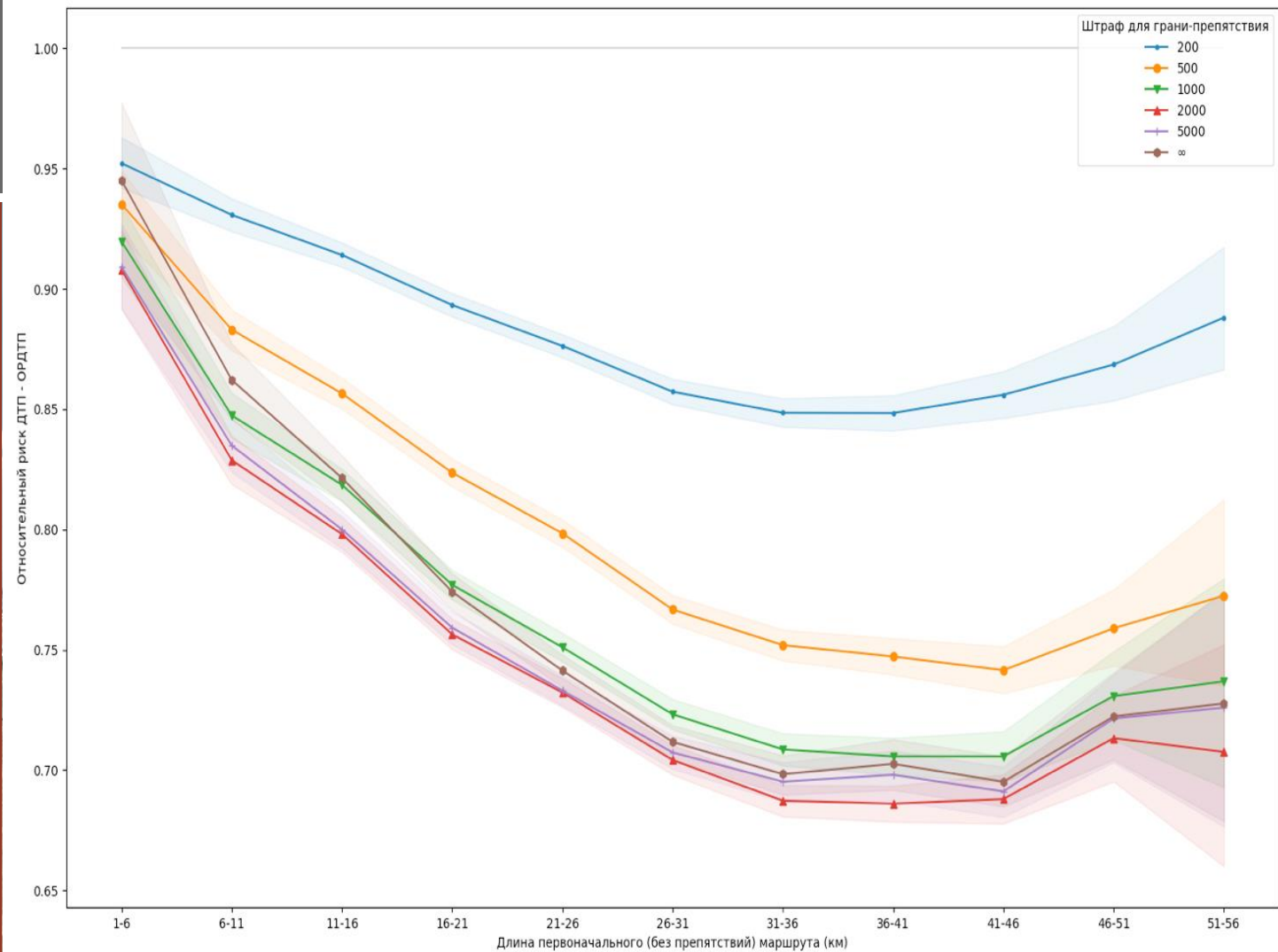
- Источником ДТП служили данные ГИБДД о 27798 ДТП, зафиксированных в Москве в 2019-2021 годах.
- *Первоначальные ДТП подверглись фильтрации*, потому что не все они (в силу, например, ошибок регистрации) принадлежали дорожной сети. Всего отфильтрованных ДТП оказалось 21956 (из 27798) (**79,0%**). Для каждого такого ДТП определялось ребро, которому оно принадлежит.
- Для статистических испытаний использовался программный пакет **SANET**, *позволяющий получить миллионы точек, равномерно распределенных вдоль дорожной сети.*
- *Испытания состояли в том, что формируются группы из 21956 точек (столько же, сколько произошло ДТП за 2019-2021 годы) и для каждой группы вновь перебираются все ребра и находится число точек, принадлежащих каждому ребру.*
- Произведя, скажем, 1000 испытаний (так и было в реальности), получим статистику ДТП в случае, когда ДТП распределены равномерно по дорожной сети.
- *Если число реальных ДТП ребра превысит 95-й перцентиль*, полученный в результате статистических испытаний, *будем считать, что число ДТП данного ребра достоверно на уровне 5%.*
- *На каждое выделенное ребро-препятствие накладывался «штраф» — к атрибуту длины ребра прибавлялось число, одно и то же для всех ребер*, что служило сигналом для алгоритма маршрутизации по возможности изменить маршрут.





## РИС. 7. Относительный риск ДТП для Москвы в зависимости от длины исходного маршрута и штрафа для ребра-препятствия с 95% доверительными интервалами .

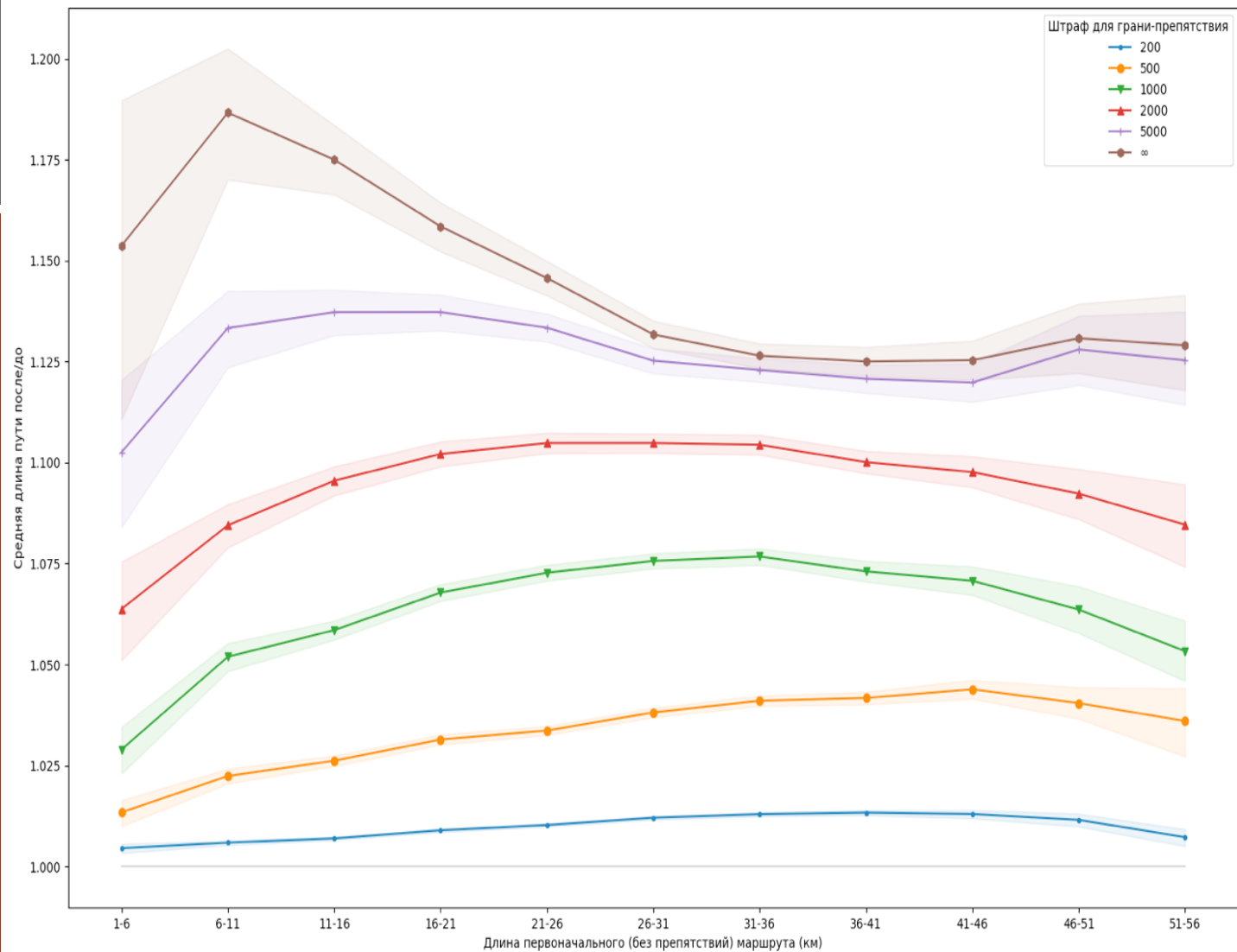
Показана зависимость относительного риска ДТП  
от длины первоначального маршрута  
при различных значениях штрафа (от 200м до бесконечности)  
Под бесконечностью понимается число, на много порядков пре-  
вышающее длину любого маршрута в Москве.





**РИС. 8. Зависимость средних отношений длин маршрутов для Москвы (Длина при обходе УПО/ Первоначальная длина) для различной длины оригинального маршрута и различных значений штрафа для ребра-препятствия**

Показана зависимость средних отношений длин маршрутов от длины первоначального маршрута.



## 9. Результаты анализа дорожной сети Москвы

Сравнение рисунков 7 и 8 показывает, что *оптимальным значением штрафа для Москвы следует считать 2000м* – при этом значении штрафа *относительный риск ДТП минимален* и достигает минимального значения 0.686, меняясь в зависимости от длины оригинального маршрут в пределах от 0.9 до 0.686.

- В графиках интересно также то, что бесконечное значение штрафа ( $10^{13}$ ) приводит и для Москвы к существенному увеличению отношения длин маршрутов на маршрутах малой длины — по той же очевидной причине — трудности прокладки маршрута при полной непроходимости одного из ребер оригинального маршрута.

Штрафы : 200, 500, 1000, 2000, 5000 и  $10^{13}$  м.

### **Выводы из Рис. 7, 8:**

*Обход препятствий, выделенных на дорожном графе г. Москвы позволяет (при значении штрафа 2000м) снизить относительный риск ДТП на 9...31 % (в зависимости от длины исходного маршрута) за счет увеличения средней длины маршрута на 6...11 % и увеличения среднего числа проходимых вершин дорожного графа на 6...28 %.*



## 10. Научная новизна

1. Автором диссертации впервые предложен интегральный **критерий качества маршрутизации**, которым служит **относительный риск ДТП**, (*ОРДТП*), равный *среднему отношению числа ДТП вдоль модифицированного* (с учетом препятствий на дорогах) *маршрута к числу ДТП вдоль эталонного* (не учитывающего препятствия) *маршрута*. Если среднее значение *ОРДТП* *значимо меньше единицы*, будем считать, что *модифицированный маршрут безопасней оригинального*.
2. *Статистически значимые кластеры ДТП используются как препятствия при построении альтернативного (более безопасного) маршрута*.
3. В качестве препятствий используются *отдельные сегменты дороги* (ребра дорожного графа), *содержащие* число ДТП, *статистически превышающее число ДТП*, полученное из предположения о равномерности распределения ДТП по дорожной сети.
4. Метод построения более безопасного маршрута, впервые предложенный автором, состоит в том, что каждое препятствие подвергается фиксированному *штрафу*, т.е. к атрибуту длины соответствующего ребра прибавляется одно и то же число. Далее путем расчета маршрутов по модифицированному и эталонному (не измененному) графу определяется *оптимальное значение штрафа, которое обеспечивает минимальное значение ОРДТП при минимальном проигрыше в длине маршрута и числе проходимых вершин дорожного графа*.

## 11. Валидация метода на дорожной сети СПб

- Только что описанные расчеты были без каких-либо изменений применены к дорожной сети Санкт-Петербурга.
- Оказалось, что *штраф 2000м* приводит к существенному снижению *относительного риска (13,9 – 36%)* ценой *небольшого увеличения средней длины маршрута (8.8-14.8%)*. Следовательно, *алгоритм с заранее определенным значением штрафа был проверен на новых данных и продемонстрировал, что он дает отличные результаты.*
- Однако, *оптимальным для Санкт-Петербурга следует признать штраф 1000м*, приводящий к *меньшему увеличению длины пути, в то время как относительный риск при этом практически не меняется* (средняя длина маршрута возрастает на 8,0-10,0% (1000м) сравнительно с 8.8-14.8% для 2000м). При этом средний относительный риск оказывается в интервале 14,5 – 36% (1000м) по сравнению с интервалом 13,9 – 36% для штрафа 2000м

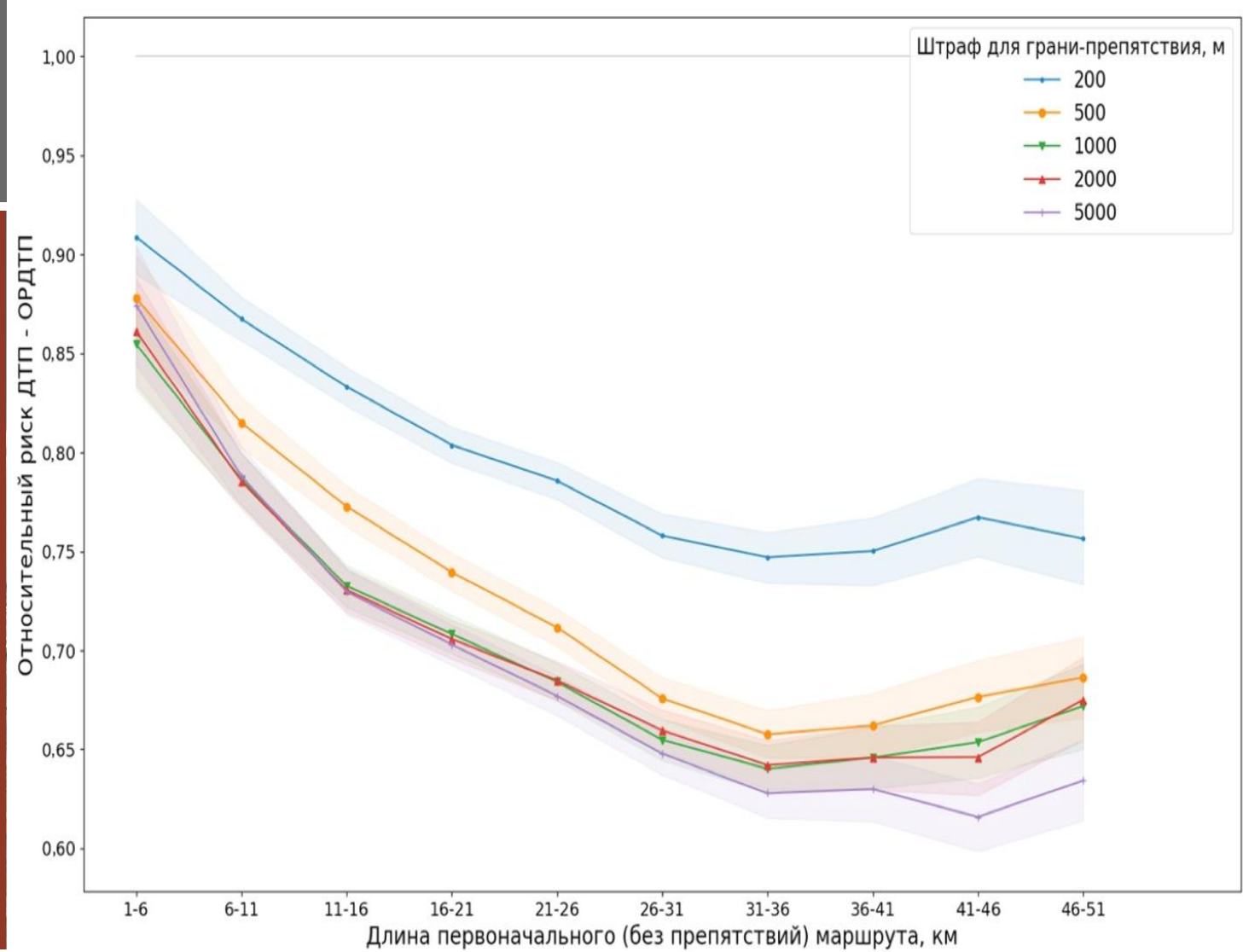
РИС.9:

ДОРОЖНАЯ СЕТЬ СПб С РЕБРАМИ-ПРЕПЯТСТВИЯМИ



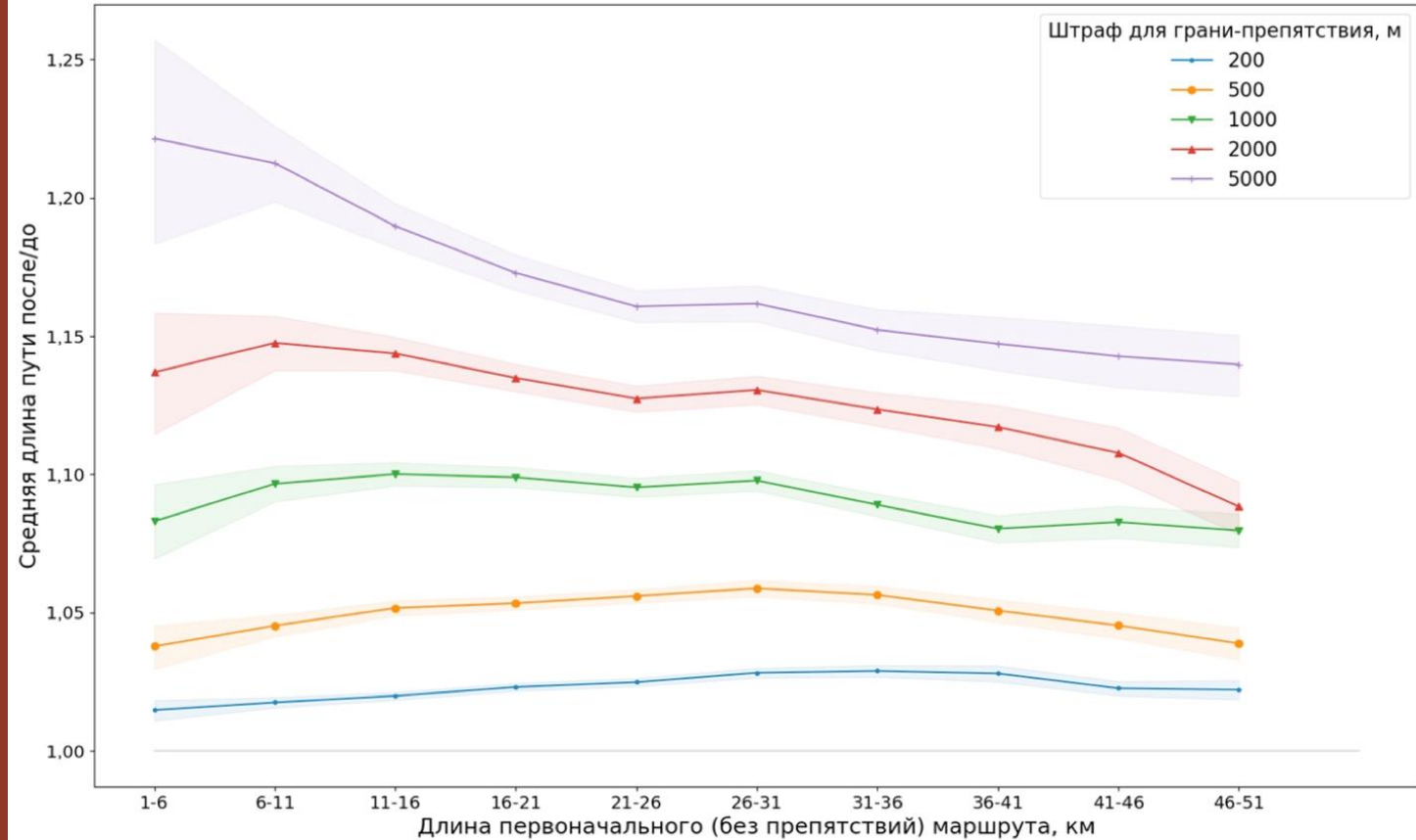


**РИС. 10. Относительный риск ДТП для Санкт-Петербурга в зависимости от длины исходного маршрута и штрафа для ребра-препятствия с 95% доверительными интервалами .**





**РИС. 11. Зависимость средних отношений длин маршрутов для Санкт-Петербурга (Длина при обходе УПО/ Первоначальная длина) для различной длины оригинального маршрута и различных значений штрафа для ребра-препятствия**





## 12. Положения, выносимые на защиту

1. Разработанный автором пакет программ путем сравнения числа реальных ДТП с числом, ожидаемым в соответствии с равномерным распределением ДТП по всей дорожной сети (с использованием имитационного моделирования Монте-Карло), позволяет выделить дорожные сегменты со статистически значимым высоким числом ДТП. Конкретные вычисления были проведены для Москвы и Санкт-Петербурга.
2. Пакет программ, разработанный автором, позволяет определить оптимальное значение штрафа, накладываемого на ребро дорожного графа, содержащее статистически значимое число ДТП. Для г. Москвы штраф равен 2000м. Это значение приводит в существенному снижению относительного риска ДТП на 9-31% за счет увеличения средней длины маршрута на 6-11%.
3. Вычисления, первоначально проведенные с помощью разработанного автором пакета программ для г. Москвы (пп. 1-2), был проверены на дорожной сети Санкт-Петербурга. Оптимальным для Санкт-Петербурга следует признать штраф 1000м, приводящий к меньшему увеличению длины пути по сравнению со штрафом 2000м, в то время как относительный риск при этом практически не меняется (средняя длина маршрута возрастает на 8,0-10,0% (1000м) сравнительно с 8,8-14,8% для 2000м). При этом средний относительный риск оказывается в интервале 14,5 – 36% (1000м) по сравнению с интервалом 13,9 – 36% для штрафа 2000м.

Штраф	относительный риск	увеличению длины пути
1000 м	- 14,5 до – 36 %	+ 8,0-10,0%
2000 м	- 13,9 до – 36 %	+ 8,8-14,8%

4. Алгоритм маршрутизации, первоначально примененный в разработанном автором пакете программ для Москвы, является устойчивым, то есть, его без каких-либо изменений (за исключением выбора оптимального штрафа) можно применять к дорожной сети других городов.

## **13. Пакет программ, модифицирующих дорожный граф в соответствии с числом ДТП, принадлежащих отдельным ребрам.**

1. Получение дорожного графа в формате OSMNX и (отдельно) ребер и вершин дорожной сети в формате ShapeFile (Приложение 1)
2. Размещение ДТП на дорожном графе. ДТП принадлежит ближайшему ребру дорожного графа (Приложение 2)
3. Подсчет числа ДТП для каждого ребра дорожного графа (Приложение 4)
4. Статистические испытания для определения ребер со статистически значимым числом ДТП. (Приложение 5)
5. Выделение ребер со статистически значимым числом ДТП (Приложение 7)
6. Создание квадратной решетки для наложения на дорожный граф (Приложение 8)
7. Построение маршрутов на модифицированном графе и сбор статистики относительного риска ДТП (Приложение 9)
8. Окончательная обработка и визуализация данных (Приложение 10)

Вспомогательные программы:

1. Оценка точности вычисления расстояний в UTM координатах с помощью формулы Евклида (Приложение 3).
2. Представление равномерно распределенных точек, полученных с помощью программы SANET, в виде .csv файла (Приложение 6).

Спасибо за  
внимание!

---

