Министерство образования и науки Российской Федерации
Московский физико-технический институт (государственный университет)
Факультет управления и прикладной математики
Кафедра «Интеллектуальные системы»
при Вычислительном центре им. А. А. Дородницына РАН
Бишук Антон Юрьевич

Применение активного обучения к графовым моделям на примере оценки рисков распространения эпидемии

010900 — Прикладные математика и физика

Бакалаврская диссертация

Научный руководитель:
к.ф.-м.н.
Зухба Анастасия Викторовна

Москва
2021 г.

Содержание

1 Введение **4**

2 Введение в курс дела **4**

2.1 Общепринятые обозначения 4

2.2 Модели (SIR, SI, SEIR(S),TG-SEIRS) 4

2.3 Разрастании эпидемии 7

2.4 Активное обучение и теория информации 7

3 Основной раздел работы **8**

3.1 Моделирование с SEIR+ 8

3.1.1 Распространение на графе 8

3.1.2 Теорема об эквивалентности людей 8

3.2 Ограничительные меры и их моделирование 10

3.3 Пагубное влияние локдауна 11

3.4 Энтропия и тестирование людей 11

3.4.1 Энтропия графа 11

3.4.2 Случай знания прошлого 12

3.4.3 Случай распределения вероятностей 12

3.4.4 Смешанный случай 13

3.4.5 Алгоритм выбора с точки зрения влияния на одну вершину . . . 13

3.4.6 Алгоритм выбора 17

4 Вычислительные эксперименты **18**

4.0.1 Пагубный эффект локдауна 18

4.0.2 Влияние ограничительных мер на поведение эпидемии 18

4.1 Уменьшение энтропии 19

5 Заключение **19**

Аннотация

Данная работа посвящена исследованию распространения эпидемии в графовых структурах. Анализируется влияние ограничительных мер на поведение эпидемии, приводится алгоритм нахождения вершин, максимально уменьшающих энтропию всей системы. Моделируется распространение эпидемии и анализируется поведение эпидемии при инициализации больными вершин с разными свойствами.

1 Введение

Эпидемия вызванная новой коронавирусной инфекцией заставляет пересмотреть отношение к бла-бла-бла. Ясно, что данная инфекция не последняя, а потому необходимо оценивать раст эпидемии и факторы, которые могут на этот рост повлиять. Однако закрытие людей на продолжительный срок на карантине приводит как к экономическим проблемам государства, так и психологическим проблем отдельных индивидуумов. В связи с этим необходимо сдерживать эпидемию средствами, которые не влекут за собой продолжительную изоляцию всего населения. Для этого необходимо понимать какие ограничительные меры в наибольшей степени влияют на поведение эпидемии. А так же понимать, какие люди в обществе имеют наибольший риск быть больными.

Сперва мы анализируем работы из прошлого и приводим теорему о разрастании и затухании эпидемии, далее мы приводим теорему об эквивалентности в локальном и глобальном смысле вершин. Затем проводится экскурс в моделирование эпидемии на основе модели SEIRS. Позже приводится анализ ограничительных мер (тестирование и изоляция, вакцинация и локдаун) на графах и их моделирование, доказательство того, что локдаун может вызывать разрастание эпидемии, приведение эксперимента, доказывающего это. Затем вводится модель пересчета рисков заражения и алгоритм нахождения вершины, наиболее сильно уменьшающей энтропию на следующей итерации. А после этого мы переносим результаты на задачи из других областей.

2 Введение в курс дела

2.1 Общепринятые обозначения

$\beta, \gamma, \sigma, \xi$ и другие принятые вещи

2.2 Модели (SIR, SI, SEIR(S),TG-SEIRS)

РЭРѳСЃСЂРчР,,РэРч/SEIRS.png

Рассматривается модель SEIRS, но для графа и потому в нашей модели учитывается эта специфика через переход из S в E.

Параметры σ, γ, ξ – остаются такими же как в классической модели и равны константным значениям.

В уравнении SEIRS выделим красным те части, которые не остаются такими же для нашей модели.

$$\begin{cases} S' = -\beta IS, \\ E' = \beta IS - \sigma E, \\ I' = \sigma E - \gamma I, \\ R' = \gamma I - \xi R. \end{cases}$$

В нашей модели будет следующая замена:

$$\beta IS \to j(\beta_{vac}, I, S, G),$$

где β_{vac} – зараженность (характеристика самого заболевания), G – характеристика графа (контактные вершины, длина рёбер и т.д.)

Система рассматривается в моменты времени $T = \{t_1, t_2, \dots t_k, \dots\}$

У нас рассматривается граф $G(V, E)$. Тогда введем следующие обозначения:

Пусть $A \in V$, тогда $S_t(A), E_t(A), I_t(A), R_t(A)$ – назовем множества вершин из A, находящиеся в момент времени t в состояниях Susceptible, Exposed, Infectious, Recovered соответственно.

Назовем $V_i = \{v_j | e_{ij} > 0\}$ – множество контактных с v_i вершин.

Введем род пояснений и одно обозначение:

$$\begin{cases} \sigma = \mathbb{P}(v_i \in I_{t+1} | v_i \in E_t), \\ \gamma = \mathbb{P}(v_i \in R_{t+1} | v_i \in I_t), \\ \sigma = \mathbb{P}(v_i \in E_{t+1} | v_i \in R_t), \\ \beta_t^i := \mathbb{P}(v_i \in E_{t+1} | v_i \in S_t). \end{cases}$$

Пусть у нас есть ряд условий взаимодействий вершин. Пусть условия выражаются в множестве параметров G, тогда пусть у нас есть L различных условий взаимодействия, тогда:

$$\beta_t^i = \sum_{l=1}^{L} |I_t(V_i^l)| \frac{1}{|V_i^l|} B_i^l$$

В частном случае, если у нас рассматривается случай, когда есть только два взаимодействия: дом (H) и работа (W), то имеем:

$$\beta_t^i = |I_t(V_i^H)| \frac{1}{|V_i^H|} B_i^H + |I_t(V_i^W)| \frac{1}{|V_i^W|} B_i^W$$

Для упрощения записей введем следующее обозначение:

$$S_t := S_t(V)$$

Необходимо узнать какие число людей в среднем переходит из одного состояние в другое за один момент времени:

$$\begin{cases} E(|R_{t+1}| - |R_t|) = E(\gamma|I_t| - \xi|R_t|), \\ E(|I_{t+1}| - |I_t|) = E(\sigma|E_t| - \gamma|I_t|), \\ E(|E_{t+1}| - |E_t|) = E(S \to E) - E(\sigma|E_t|), \\ E(|S_{t+1}| - |S_t|) = E(\xi|R_t|) - E(S \to E). \end{cases}$$

Здесь $E(S \to E)$ – математическое ожидание людей, переходящих из состояния S в состояние E в момент времени t.

Если β_t^i – считать вероятностью, то имеем:

$$E(S \to E) = E(\sum_{v_i \in S_t} \beta_t^i)$$

Теперь если мы предполагаем, что на k-ой итерации мы знаем всё о системе, но хотим узнать что станет с системой на $k+1$ итерации, то система переписывается в следующем виде:

$$\begin{cases} E(|R_{t+1}|) - |R_t| = \gamma|I_t| - \xi|R_t|, \\ E(|I_{t+1}|) - |I_t| = \sigma|E_t| - \gamma|I_t|, \\ E(|E_{t+1}|) - |E_t| = \sum\limits_{i:v_i \in S_t} |I_t(V_i^H)| \frac{1}{|V_i^H|} B_i^H + \sum\limits_{i:v_i \in S_t} |I_t(V_i^W)| \frac{1}{|V_i^W|} B_i^W - \sigma|E_t|, \\ E(|S_{t+1}|) - |S_t| = \xi|R_t| - \sum\limits_{i:v_i \in S_t} |I_t(V_i^H)| \frac{1}{|V_i^H|} B_i^H - \sum\limits_{i:v_i \in S_t} |I_t(V_i^W)| \frac{1}{|V_i^W|} B_i^W. \end{cases}$$

Если же мы не знаем точное состояние системы на прошлой итерации, а знаем лишь вероятности вершин принадлежать к тому или иному множеству, тогда система переписывается в следующем виде:

$$\begin{cases} E(|R_{t+1}| - |R_t|) = E(\gamma|I_t| - \xi|R_t|), \\ E(|I_{t+1}| - |I_t|) = E(\sigma|E_t| - \gamma|I_t|), \\ E(|E_{t+1}| - |E_t|) = \sum\limits_{i:v_i \in V} \frac{B_i^H}{|V_i^H|} E\left(|I_t(V_i^H)| \cdot \mathbb{P}(v_i \in S_t)\right) + \\ \qquad + \sum\limits_{i:v_i \in V} \frac{B_i^W}{|V_i^W|} E\left(|I_t(V_i^W)| \cdot \mathbb{P}(v_i \in S_t)\right) - E(\sigma|E_t|), \\ E(|S_{t+1}| - |S_t|) = E(\xi|R_t|) - \sum\limits_{i:v_i \in V} \frac{B_i^H}{|V_i^H|} E\left(|I_t(V_i^H)| \cdot \mathbb{P}(v_i \in S_t)\right) - \\ \qquad - \sum\limits_{i:v_i \in V} \frac{B_i^W}{|V_i^W|} E\left(|I_t(V_i^W)| \cdot \mathbb{P}(v_i \in S_t)\right). \end{cases}$$

2.3 Разрастании эпидемии

Определение 2.1. *Критический уровень заражения (эпидемиологический порог) λ_c – значение уровня заражения, такое, что при значении λ меньше него, эпидемия затухает, а при значении λ больше λ_c, эпидемия разрастается.*

Определение 2.2. *Средняя связность $\langle k \rangle$ – величина, равная $\sum\limits_k kP(k)$, где $P(k)$ – распределение связности в графе.*

Теорема 2.1. *В модели SIR критическое значение уровня заражения выражается формулой:*

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

2.4 Активное обучение и теория информации

Слова про то, что такое активное обучение и с чем его едят + написать причем тут теория информации и энтропия

3 Основной раздел работы

3.1 Моделирование с SEIR+

Для моделирвоания эпидемии использовал библиотека SEIR+ моделированная под наши нужды:)

3.1.1 Распространение на графе

3.1.2 Теорема об эквивалентности людей

Определение 3.1. *Назовем вектор x – вектором рисков, представляющий из себя вектор длины 2^n, где n – число вершин в графе, строящийся по следующему правилу: на первом месте стоит вероятность того, что никто не более, далее вероятность, что более только первая вершина, ..., болеют первые две вершины, ..., болеют все вершины.*

Определение 3.2. *Назовем матрицу A – матрицей распространения эпидемии, представляющую из себя матрицу размера $2^n \times 2^n$, строящуюся по определенным правилам: в строках находятся состояния системы на текущей итерации, в столбцах состояния, в которые возможно перейти, в ячейках матрицы – вероятность такого перехода.*

Определение 3.3. *A^k - представление эпидемии на k-ом шаге. Если взять за x - вектор риска в начальный момент времени, то $x^T A$ - вектор рисков на k-ой итерации.*

Определение 3.4. *Назовем «схлопыванием» матрицы A операцию $A^k_{2^n \times 2^n} \rightarrow \overline{A}^k_{2^n \times (n+1)}$, осуществляемую умножением матрицы A на матрицу F, где $\overline{A}^k$ – матрица, у которой «схлопнулись» те столбцы, которые соответствуют одному и тому же числу больных.*

Определение 3.5. *Если в матрице A^k из одной строки можно получить другую путем перенумерации вершин графа, то такие вершины будет называть эквивалентными в глобальном смысле.*

Определение 3.6. *С точки зрения числа больных удобно рассматривать матрицу $\overline{A}^k$. Тогда строки, которые совпадают в «схлопнутой» матрице будем называть эквивалентными в локальном смысле.*

Определение 3.7. *Будем называть размером окрестности степени k вершины v – число вершин, до которых можно добраться следуя по ребрам исходного графа за k переходов.*

Число заболевших в графе на k-ом шаге можно оценить сверху как размер окрестности степени k.

Если первые окрестности вершин v_1 и v_2 совпадут, то строки матрицы $\overline{A}$, соответствующие случая, когда болеет только v_1 и только v_2 так же совпадут, т.е. будут эквивалентны в локальном смысле.

Теорема 3.1. *Пусть дан граф $G(V, E)$, пусть в нем n вершин заражено и $n - V$ здоровых, тогда число заболевших на следующей итерации монотонно от величины окрестности для всех больных вершин.*

Proof.

Для простоты выкладок возьмем $n = 1$ и примем степень этой вершины за k.

Поскольку нам известна структура графа, то мы можем составить матрицу перехода системы A.

В строке, соответствующей нашей ситуации, где более один конкретный человек, число ненулевых элементов равно $C_{k+1}^1 + C_{k+1}^2 + ... + C_{k+1}^{k+1}$.

«Схлопним» матрицу A до матрицы $\overline{A}$.

Тогда в матрице $\overline{A}$ в строке, соответствующей инициализированному заболевшему, в первом столбце будет стоять одно число, означающее вероятность того, что никто не заболел, а текущий человек выздоровел, во втором столбце будет сумма из C_{k+1}^1 чисел, которая означает вероятность того, что заболел один человек, т.д. и в $k+1$-ом столбце будет стоять вероятность того, что заболело k человек и текущий не выздоровел. Во всех дальнейших клетках будут стоять нули.

И так, мы получили распределение числа заболевших после первой итерации. Однако, если теперь мы будем рассматривать не $\overline{A}$, а матрицу $\overline{A}^m$, полученную путем «схлопывания» матрицы A^m, то продевал те же действия, что и выше, мы получим распределение числа заболевших на m-ой итерации.

Поскольку мы получили распределение, то мы можем посчитать математическое ожидание числа заболевших на m-ой итерации. Для этого умножим строку матрицы $\overline{A}^m$ на вектор $w = [0, 1, 2, ..., |V|]$ (будем считать $w_i = i - 1$), представляющий из себя вектор из числа заболевших.

Обозначим число заболевших на m-ой итерации за I_m. Тогда:

$$E(I_0) = \sum_{i=1}^{k}(i-1) \cdot \sum_{j=1}^{C_{k+1}^{i}} p_{i,j}$$

Здесь $p_{i,j}$ – вероятности перехода в матрице A в группе i (группа формируется по количеству заболевших) по номеру ненулевого элемента j.

В общем случае вместо k необходимо использовать k_m – размер окрестности m-го порядка. А за $p_{i,j}$ – вероятности перехода в матрице A^k в группе i (группа формируется по количеству заболевших) по номеру ненулевого элемента j.

3.2 Ограничительные меры и их моделирование

Для борьбы с распространением эпидемии используется ряд противоэпидемиологических мер:

- Вакцинация
- Изоляция больных
- Введение lockdown

И если первые два способа сдерживания эпидемии в худшем случае может не помочь, то в случае локдауна, ограничение может отрицательно повлиять на сдерживание эпидемии.

Изоляция вершин происходит после проверки людей тестом, т.е. подразумевает ранжирование людей по вероятности того, что человек болен.

Моделирование вакцинации происходит путём замены статуса человека с S на R.

Изоляции больных моделируется путём удаления вершины из графа и всех инцидентных с ней рёбер.

Моделирование локдауна происходит путём замены имеющегося графа на граф, состоящий из клик небольшого размера.

Лемма 3.1. *Ограничительная мера тестирование и изоляция отрицательно скорелированна с прогнозируемым числом заболевших.*

Proof:

- *Пусть была протестирована вершина и она оказалась здоровой, то это значит, что вершину изолировать не нужно, поскольку она не сможет никого заразить, но может заразиться, что увеличит число заболевших. Но это бы произошло с той же вероятность, если бы мы не проверяли данную вершину.*

- *Пусть мы проверили вершину (пусть v_i) и она оказалась больной. Если эта вершина не будет изолирована, то верхней оценкой прироста числа заболевших от этой вершины будет $\sum\limits_{v_j \in V_i} \beta$. А если эта вершина будет изолирована, то верхней оценкой будет 0.*

Получили, что тестирование и изоляция больных не ухудшает эпидемиологическую обстановку.

Лемма 3.2. *Ограничительная мера вакцинация отрицательно скорелированна с прогнозируемым числом заболевших.*

Proof:

- *Пусть мы выбрали вершину и провакцинировали её. Тогда если она в будущем будет изолированной, то вакцинация никак не повлияет на прирост числа зараженных.*

- *Пусть вершина не изолированная и имеет больные вершины, тогда вакцинация гарантированно не увеличит число заболевших на следующих итерациях по сравнению с тем, если бы мы не вакцинировали*

Получили, что вакцинирование не ухудшает эпидемиологическую обстановку.

3.3 Пагубное влияние локдауна

Теорема 3.2. *Теорема*

3.4 Энтропия и тестирование людей

3.4.1 Энтропия графа

Теперь будем считать, что граф это систему, в вершинах, которой стоят вероятности того, что вершина имеет метку 1 (вершина больна). Иными словами чем меньше вероятность находится в вершине, тем выше вероятность того, что в вершине находится метка 0.

Тогда под энтропией будем понимать следующую величину:

$$H(\{x_i\}_{i=1}^n) = -\left(\sum_{i=1}^n p_i \log(p_i) + \sum_{i=1}^n (1-p_i)\log(1-p_i)\right)$$

Максимум достигается в точке $x_{opt} = \frac{1}{2}$, это значит, что проверка человека, вероятность которого ближе всего к x_{opt} максимально уменьшит энтропию системы, получаемую от одного слагаемого.

3.4.2 Случай знания прошлого

Известна вся информация о системе на прошлой итерации.

1. Если i-ая вершина была здорова и не имела контактов с больными, то $p_i^{t-1} = p_i^t = 0$.

2. Если i-ая вершина была больна, то $p_i^t = 1 - \gamma$.

3. Если i-ая вершина была здорова, но контактировала с $\{x_{i_k}\}_{i_k}^K$ больными вершинами, то $p_i = f(\beta, K)$ (или в общем случае $f(\{\beta_l\}, \{\theta_l\})$ $f(\beta, k)$:

   a) β;

   b) $\beta + (1-\beta)\sum\limits_{l=1}^{K-1}(\frac{1}{2})^l$;

   c) $\beta + (1-\beta)e^{-c/(K-1)}$;

   d) В общем случае: $\beta + (1-\beta)\sum\limits_{l=1}^{K-1} g(l)$,

$$g(l) : \begin{cases} \sum\limits_{l=1}^{K-1} g(l) \to 1 & , K \to 1 \\ g(m) > g(m+1) \end{cases}$$

3.4.3 Случай распределения вероятностей

Известно распределение вероятностей на прошлой итерации $\{p_i^{t-1}\}_{i=1}^n$. При этом считаем, что β – это вероятность заразиться от больного, то есть β – верхняя оценка вероятности заразиться:

1. если $\exists i'$: $p_{i'}^{t-1} = 0$, $\forall j : \exists (i'j) : p_j^{t-1} = 0 \hookrightarrow p_{i'}^t = 0$

2. $p_i^t = p_i^{t-1}(1-\gamma) + (1-p_i^{t-1})f(\{p_{i_k}\}, \beta)$
   $f(\beta, \{p_{i_k}\}_{k=1}^K)$:

a) Наивный вариант $\frac{\beta}{K}\sum\limits_{k=1}^{K} p_{i_k}$;

b) $f(\beta, \{p_{i_k}\}_{k=1}^{K}) = \beta e^{-c/\sum\limits_{i_k} p_{i_k}}$;

c) В общем случае:

$$f(\beta, \{p_{i_k}\}_{k=1}^{K}) : \begin{cases} f(\beta, \{p_{i_k}\}_{k=1}^{K-1}) < f(\beta, \{p_{i_k}\}_{k=1}^{K}) \\ f(\beta, \{p_{i_k}\}_{k=1}^{K}) \to \beta \\ \forall n, \forall p_{i_k} \hookrightarrow f(\beta, \{p_{i_k}\}_{k=1}^{K}) < \beta \end{cases}$$

Известно распределение вероятностей на прошлой итерации $\{p_i^{t-1}\}_{i=1}^{n}$. И у части вершин мы знаем их точную метку. В силу построения модели, метки можно воспринимать как вероятности, поэтому для этого случая верны формулы из предыдущего пункта.

3.4.4 Смешанный случай

Известно распределение вероятностей на прошлой итерации $\{p_i^{t-1}\}_{i=1}^{n}$. И у части вершин мы знаем их точную метку. В силу построения модели, метки можно воспринимать как вероятности, поэтому для этого случая верны формулы из предыдущего пункта.

3.4.5 Алгоритм выбора с точки зрения влияния на одну вершину

Необходимо найти такую вершину x_i, изменение вероятности p_i^{t-1} которой путем тестирования максимально приблизит p_j^t какой-то из вершин к $\frac{1}{2}$. Более формально:

Пусть p_i^t – вероятность того, что в i-ой вершине находится метка 1 до того, как мы кто-то был протестирован.

$\tilde{p}_i^t$ – вероятность того, что в i-ой вершине находится метка 1 после того, как кто-то был протестирован.

Тогда необходимо найти такую вершину $x_{i'}$: $\forall i \hookrightarrow |p_i^t - \frac{1}{2}| > |\tilde{p}_{i'}^t - \frac{1}{2}|$ и для $\forall i \neq i' \hookrightarrow |\tilde{p}_i^t - \frac{1}{2}| > |\tilde{p}_{i'}^t - \frac{1}{2}|$

Теперь легко заметить, что для наивного пересчета вероятностей ($p_i^t = p_i^{t-1}(1-\gamma) + (1-p_i^{t-1})\frac{\beta}{K}\sum\limits_{k=1}^{K} p_{i_k}^{t-1}$) выполнено $E_j \tilde{p}_i^t = p_i^t, \ \ \forall j$, что значит в такой формулировке нет смысла проверять кого-то с прошлой итерации.

Тогда если $p_i^t = p_i^{t-1}(1-\gamma) + (1-p_i^{t-1})\beta e^{-c/\sum\limits_{i_k} p_{i_k}^{t-1}}$, то

$$\Delta p_i^t = \begin{cases} (1-p_i^{t-1})\beta\left(-e^{-c/\sum\limits_{i_k} p_{i_k}^{t-1}} + e^{-c/(\sum\limits_{i_k} p_{i_k}^{t-1}+1-p_j^{t-1})}\right) & \text{, с вероятностью } p_j^{t-1} \\ (1-p_i^{t-1})\beta\left(-e^{-c/\sum\limits_{i_k} p_{i_k}^{t-1}} + e^{-c/(\sum\limits_{i_k} p_{i_k}^{t-1}-p_j^{t-1})}\right) & \text{, с вероятностью } 1-p_j^{t-1} \end{cases}$$

Тогда:

$$E_j\Delta p_i^t = (1-p_i^{t-1})\beta\left(-e^{-c/\sum\limits_{i_k} p_{i_k}^{t-1}} + (1-p_j^{t-1})e^{-c/(\sum\limits_{i_k} p_{i_k}^{t-1}-p_j^{t-1})} + p_j^{t-1}e^{-c/(\sum\limits_{i_k} p_{i_k}^{t-1}+1-p_j^{t-1})}\right)$$

И тогда: $\tilde{p}_i^t = p_i^t + E_j\Delta p_i^t$

И решаем следующую оптимизационную задачу:

$$|\tilde{p}_i^t - \frac{1}{2}| \to \min_{i,j}$$

Или если перейти к гладкой версии:

$$A = (\tilde{p}_i^t - \frac{1}{2})^2 \to \min_{i,j}$$

Для простоты записи будем считать, что $c = 1$:

$$A = \left(-\frac{1}{2} + (1-\gamma)p_i^{t-1} + \beta(1-p_i^{t-1})(q_j^{t-1}\exp\big(\frac{-1}{\sum\limits_{i_k} p_{i_k}^{t-1} + 1 - q_j^{t-1}}\big) + (1-q_j^{t-1})\exp\big(\frac{-1}{\sum\limits_{i_k} p_{i_k}^{t-1} - q_j^{t-1}}\big))\right)^2$$

Находить оптимум необходимо численно решая уравнение $A' = 0$. Тогда получим решение ${q_{opt}}_j^{t-1} = f(\sum\limits_{i_k} p_{i_k}^{t-1}, p_i^{t-1}, \gamma, \beta)$.

Поскольку данная функция меняет поведение, то минимум достигается либо в оптимуме, либо на границе, т.е. необходимо проверить три точки.

То есть оптимум либо в $q_j^{t-1} = 0$, либо $q_j^{t-1} = {q_{opt}}_j^{t-1}$, либо $q_j^{t-1} = \min(1, \sum\limits_{i_k} p_{i_k}^{t-1})$

Тогда сам алгоритм определения лучшей вершины для проверки:

1. упорядочим $\{p_j^{t-1}\}_{j=1}^n$. (за $O(n\log(n))$ если мы не можем позволить дополнительную память или за $O(n)$, если проблем с дополнительной памятью нет. !!! Очень тонкий момент! Нужно что-то придумать с тем, как делать это быстрее, иначе махание руками с производными не имеет смысла.
2. Для каждой вершины на итерации t получаем точку в которой достигается минимум $(\tilde{p}_i^t - \frac{1}{2})^2 \to \min_{i,j}$
3. Бинарным поиском ищем какая из вероятностей на прошлой итерации ближе всего к оптимуму, для неё считаем значение функции и запоминаем вершину, если функция уменьшилась
4. в конце алгоритма получим вершину, которую необходимо проверить.

Для того, чтобы не приходилось сортировать вероятности соседей создадим специальную структуру:

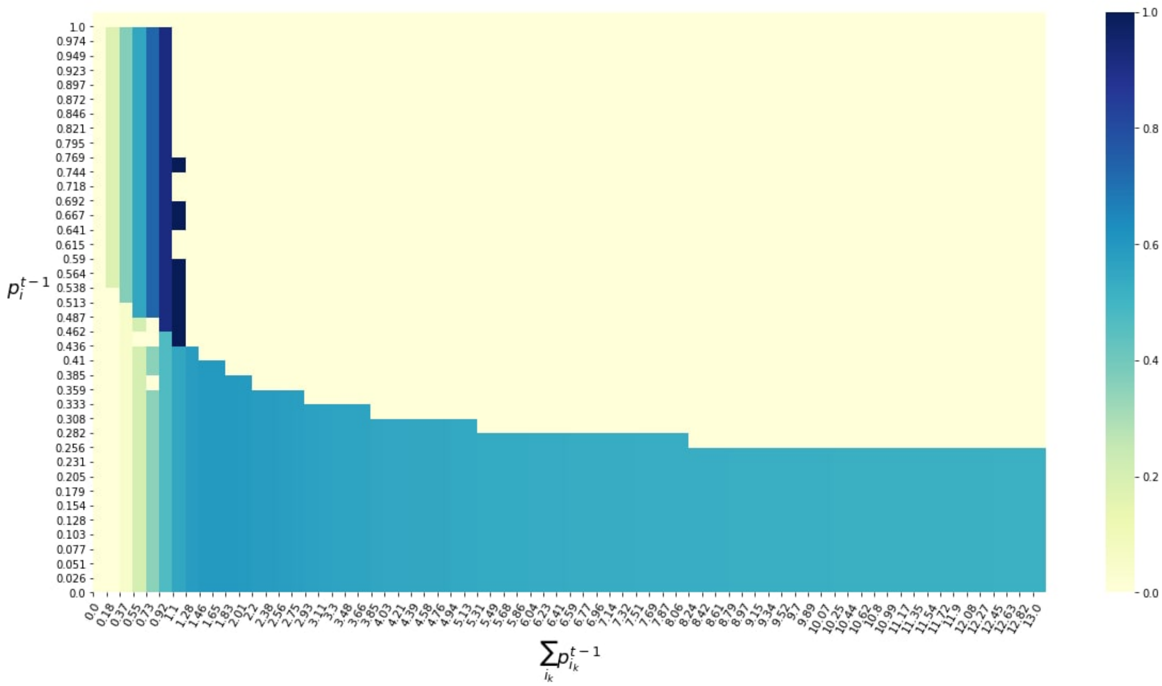
- Список вершин.
  1. Умеет получать за O(n) все вероятности в каждой вершине
  2. Умеет возвращать список вершин с вероятностями на итерации t за $O(1)$
  3. Умеет возвращать вершину на итерации t за $O(1)$.
- Вершина.
  1. Хранит в себе номер итерации, свой номер, своё значение на итерации t и $t-1$, а также список своих соседей на прошлой итерации с вероятностями в специальном виде
  2. Специальный вид получается следующим образом: Список вершин возвращает словарь $\{1 : p_1^{t-1}, \dots, n : p_n^{t-1}\}$, далее этот массив сортируется по значениям и получается список вида $[(1_s, p_{1_s}^{t-1}), \dots, (n_s, p_{n_s}^{t-1})]$, далее он преобразуется в словарь $\{1 : (p_{1_s}^{t-1}, 1_s), \dots, n : (p_{n_s}^{t-1}, n_s)\}$. Далее в каждой вершине смотрим на всех соседей и собираем информацию из словаря о соседях и формируем словарь $\{1_s : (p_{1_s}^{t-1}, 1), \dots, n_s : (p_{n_s}^{t-1}, n)\}$, и получим вероятности в отсортированном виде.
  3. Важно, чтобы всё работало быстро необходимо чтобы список вершин собирал все вершины, сортировал их sorted(array-vertix, key = lambda

x: -x.prob), проход по всем вершинам и запись в vertix.position = num_in_sort_arr

Этот алгоритм работает за $O(E+V)$.

Мы можем построить следующую карту рекомендаций проверки, где по оси x отложена суммарная вероятность метки 1 соседей, по оси y вероятность вершины на прошлой итерации.

Тогда в зависимости от этих двух вершин можно понять какую вероятность соседей необходимо проверить, чтобы уменьшить вклад энтропии для текущего слагаемого.



У этого алгоритма очевидно есть проблема, выраженная в том, что проверка вершины меняет вероятности не только для рассматриваемой вершины, но и для других вершин, смежных с проверяемой вершиной, однако показывает как можно максимизировать вклад от конкретного слагаемого.

Это например нужно в случае если мы хотим узнать какие вершины являются самыми важными для своих соседей и отранжировать их по такому принципу, что первыми идут те вершины, которые для наибольшего числа вершин являются важными

3.4.6 Алгоритм выбора

Теперь будем рассматривать другой подход. Теперь мы будем тестировать рассматриваемую вершину и смотреть как от этого поменяется энтропия всей системы за счет изменения энтропии соседних вершин.

- Проходимся по всем рёбрам и записываем в вершины вероятности соседей.

- Пересчитываем вероятности для каждой вершины на основе вероятностей с прошлой итерации.

- На основе новых вероятностей считаем H_i – вклад i-ой вершины в энтропию системы.

- Проходимся по всем рёбрам и для каждой вершины считаем величину $H_i^t + \sum\limits_{i_k} H_{i_k}^t$.

- Теперь для каждой вершины считаем, что мы проверили её, тогда посчитаем матожидание вклада в энтропию системы окрестности проверенной вершины. Она равна $p_i \sum\limits_{i_k} \tilde{H}_{i_k} + (1 - p_i) \sum\limits_{i_k} \tilde{\tilde{H}}_{i_k}$

- И тогда необходимо найти ту вершину, для которой величина

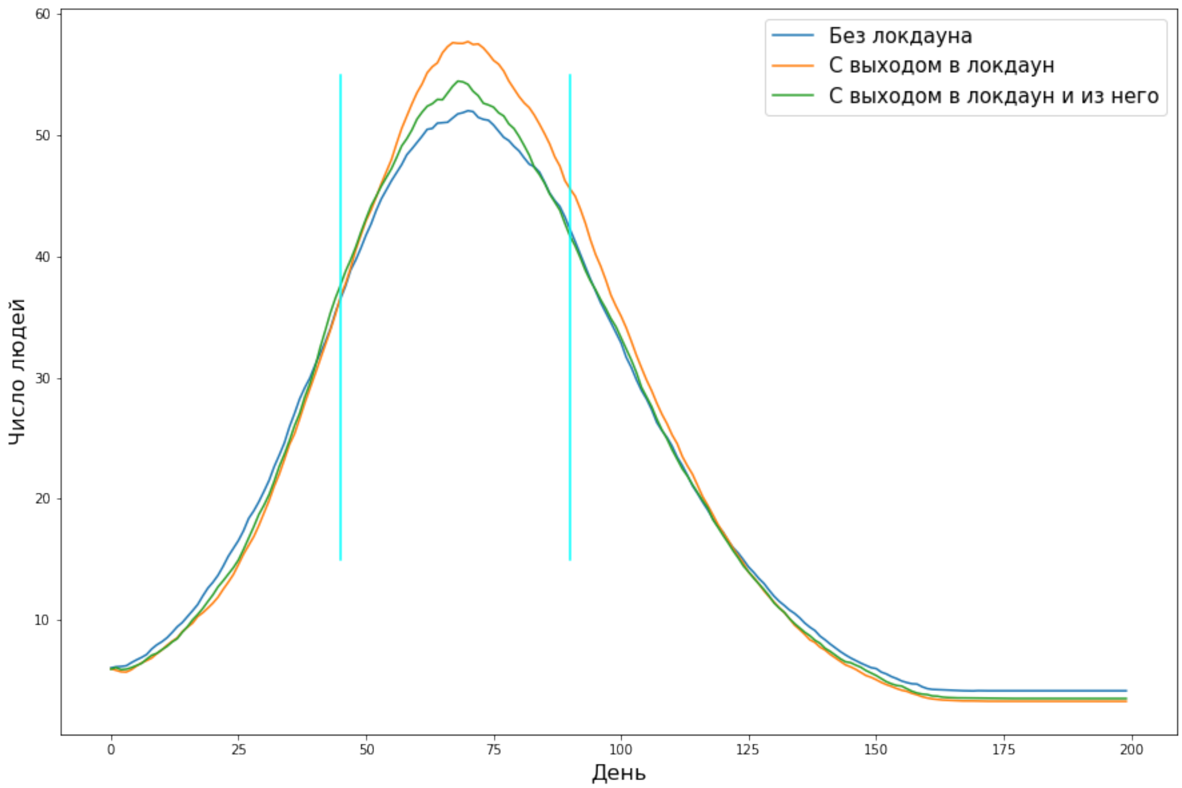
$$\left| H_i^t + \sum_{i_k} H_{i_k}^t - (p_i \sum_{i_k} \tilde{H}_{i_k} + (1 - p_i) \sum_{i_k} \tilde{\tilde{H}}_{i_k}) \right|$$

  будет максимальной, поскольку тогда мы найдем ту вершину, проверка, которой даст наибольшее уменьшение энтропии системы.

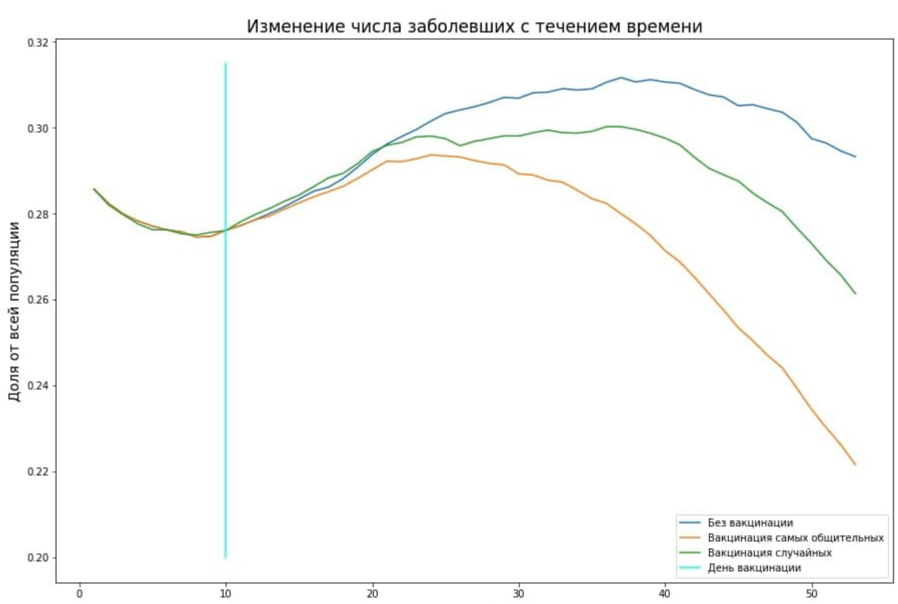
Данный алгоритм работает за $O(E + V)$.

4 Вычислительные эксперименты

4.0.1 Пагубный эффект локдауна



4.0.2 Влияние ограничительных мер на поведение эпидемии



4.1 Уменьшение энтропии

5 Заключение

Основные результаты работы

- Предложен критерий «плохого» локдауна
- Предложен алгоритм для проверки людей
- Показано влияние различных ограничительных мер на распространение эпидемии

В дальнейшем планируется исследовать другие способы агрегации вероятностей, учесть проход назад, а также явным образом использовать время контакта.