

Исследование распространения эпидемий в графовой модели SIR и влияния локдауна на динамику заболеваемости популяции

Сёмкин К.

Бишук А.

Аннотация

Задача создания и обоснования математических моделей распространения эпидемий всегда являлась и будет являться актуальной в наше время. Существуют оправданные и реалистичные модели распространения, такие как SIR и SIER, основанные на дифференциальных уравнениях динамики количества больных и здоровых жителей, но их область применимости ограничивается большими масштабами наблюдения. В данной работе рассматривается аналог SIR, основанный на графе контактов, для анализа заболеваемости на малых масштабах (предприятие, небольшая коммуна), исследуется вероятностная динамика распространения эпидемии в зависимости от параметров модели, а также влияние разных ограничительных мер. В частности интересно обоснование противоречивого эффекта, связанного с ростом заболевших при введении локдауна. Теоретические результаты демонстрируются проведением численных экспериментов посредством симулирования эпидемии на графе контактов, а также на основе реальных данных.

Введение

В связи с известными событиями интерес к эпидемиологии и её методам сильно вырос. Для понимания динамики протекания короновирусной инфекции в 2020 году в разных местах Земли и на разных масштабах, а также для подготовки к будущим вспышкам заболеваемости, как никогда актуально построение адекватных математических моделей развития болезней в людских популяциях [3]. Классические техники моделирования эпидемий опираются на параметризованные автономные системы дифференциальных уравнений, описывающие динамику изменения количества болеющих и здоровых людей. Эти модели дают хорошее понимание протекания болезни на больших масштабах (города, страны), но не способны описывать заболевание в небольших общественных структурах, например, промышленное предприятие, небольшую деревню или студенческое общежитие. В данной работе исследуется графовый подход к моделированию распространения инфекции, а именно вводится *граф контактов*, по которому болезнь может «кочевать». В качестве представления болезни используется стандартная для эпидемиологии модель SIR/SEIR [7], в которой каждому человеку (вершине в графе) сопоставляется некоторое состояние (больной, здоровый и т.д.), после чего в дискретном времени происходят смены этих состояний с некоторыми вероятностями и система эволюционирует.

Также исследуются различные эффекты от мер по борьбе с инфекцией, таких как тестирование, изоляция и, самое интересное, *локдаун*. Именно ему уделяется основное внимание, так как его введение может привести к необычному последствию — росту заболеваемости среди населения. Но обнаружить такое поведение в стандартных моделях не представляется возможным, поэтому цель данной работы — найти условия возникновения такого эффекта в модели и продемонстрировать его на численных экспериментах.

Изучение эпидемий на больших популяциях позволяет моделировать этот процесс в среднем, и даже получать точные аналитические решения [6]. В зависимости от поставленной прикладной задачи возникает необходимость моделировать процесс эпидемии с разной степенью подробности. Так, например, простейшая модель SI [1] рассматривает всего два состояния: больной и здоровый. В этой модели не рассматривается формирование иммунитета: здоровый всегда может заразиться при контакте с инфекцией. Существуют модели, рассматривающие дополнительно формирование иммунитета, инкубационный период, летальные исходы и многие другие возможные состояния. Одной из таких моделей является SEIR(S) [2]. Моделирование в среднем не подходит для небольших или слишком разнородных популяций. Эту проблему позволяют решить модели распространения эпидемии на графах [9], [10]. Распространение эпидемии на графе контактов можно рассматривать, например, при помощи цепи Маркова [5]. Однако моделирование распространения болезни на больших графах со сложной структурой имеет высокую алгоритмическую сложность. Наиболее распространенной является задача прогнозирования течения эпидемии [8] и оценка индивидуальных рисков. Результаты изучения распространения эпидемии на графах могут быть использованы не только для анализа заболеваний. Например, распространение слухов или автомобильного трафика можно описать схожим математическим аппаратом [4]. Фундаментом для данной статьи является [11], где, в частности, введена модель болезни на графе и где в её рамках исследован эффект локдауна.

В работе ставится задача обобщить модель из [11], сформулировать новые условия возникновения роста заболеваемости при введении карантина и явно показать этот эффект в численном эксперименте. Т.о. появится возможность испытывать обновлённую модель в более широком спектре реальных ситуаций, а также пересмотреть локдаун как однозначно позитивную меру противодействия эпидемии.

Постановка задачи

Формально задача состоит в выявлении зависимости роста заболеваемости при введении локдауна от графа контактов и параметров динамики развития болезни.

Пусть $G = (V, E)$ — исходный граф контактов, $G^q = (V^q, E^q)$ — граф контактов при введении карантинного режима. Два графа строятся на одних и тех же вершинах. G можно интерпретировать как контакты людей в рабочее время, граф имеет произвольную структуру, но является плотным. G^q должен представлять изоляцию вершин друг от друга, поэтому это граф состоящий из множества клик небольшого относительно $|V|$ размера.

Каждая вершина может находиться в трёх возможных состояниях: **S I R**, можно представлять это как *раскраска графа* в 3 цвета.

Ребра этих графов $w_{ij} \in [0, 1]$ и интерпретируются как доля совместного времяпрепровождения вершин (например, от 24 часов).

Болезнь параметризуется 3 параметрами:

- $\gamma = \mathcal{P}(I \rightarrow R)$ — вероятность "выздороветь"
- $\sigma = \mathcal{P}(R \rightarrow S)$ — вероятность стать снова подверженным заболеванию.
- $\beta \sim \mathcal{P}(S \rightarrow I)$ — "заразность" болезни.

Все параметры глобальные, т.е. одинаковы для всех вершин в графе. $\mathcal{P}(S \rightarrow I)$ зависит от β и от множества соседей данной вершины.

Будем понимать под G_t граф контактов на дискретном временном шаге t , т.е. его графовую структуру, а также состояния каждой вершины в данный момент. Под $\#I$ будем понимать множество больных вершин в графе (или кол-во больных в графе, в зависимости от контекста).

Т.о. задача состоит в поиске условий на t_0 , G и G_q , при которых $\max_{t \geq t_0} I(G_t^q) \geq \max_{t \geq t_0} I(G_t)$, где t_0 момент введения локдауна. Другая возможная постановка: найти условия на те же параметры, при которых $\beta_{ij}^q \geq \beta_{ij}$.

Теоретическая часть

Эпидемия как дискретная цепь Маркова

Для начала восполним одну неточность и определим правило вычисления $\mathcal{P}(S \rightarrow I)$. Для этого будем опираться на предположение, что взаимодействие вершин, а значит их заражение друг друга, происходят попарно независимо. Тогда, имея две связанные вершины u, v определяем $\mathcal{P}(u \in I_{t+1} | u \in S_{t+1}) = \mathbb{P}(v \in I_t) \cdot \beta \cdot w_{uv} \Rightarrow \mathcal{P}(u \notin I_{t+1} | u \in S_{t+1}) = 1 - \mathbb{P}(v \in I_t) \cdot \beta \cdot w_{uv}$. Если у вершины u больше соседей, то вероятность не заразиться ни от одной на следующем шаге $\mathcal{P}(u \notin I_{t+1} | u \in S_{t+1}) = \prod_{v \in N(v)} (1 - \mathbb{P}(v \in I_t) \cdot \beta \cdot w_{uv})$, здесь $N(v)$ — множество соседей для вершины. Значит, вероятность заразиться на след. шаге для здоровой вершины:

$$\mathcal{P}(u \in I_{t+1} | u \in S_{t+1}) = 1 - \prod_{v \in N(v)} (1 - \mathbb{P}(v \in I_t) \cdot \beta \cdot w_{uv})$$

Далее, структура задачи говорит сама за себя — необходимо использовать формализм марковских цепей. Можно вводить цепь для каждой вершины графа, но из-за связи с другими вершинами, эта цепь будет неоднородна. Поэтому будем вводить дискретную цепь на множестве состояний E , состоящей из всех возможных конфигураций графа G , т.е. E это всевозможные раскраски графа в три цвета. Т.о. $|E| = 3^{|V|}$ — конечное число. Более того, эта цепь будет однородной, ведь правила перехода для вершин в определённых состояниях полностью определяются параметрами эпидемии и состояниями соседей, т.е. для любой конфигурации графа определены и постоянны вероятности перехода в любую другую конфигурацию. В итоге, получаем дискретную, конечную, однородную марковскую цепь.

Думать об этой цепи в терминах конфигураций графа довольно неудобно, намного проще мыслить в каждой вершине тройку чисел $(p_S^u, p_I^u, p_R^u)^T$ — распределение вероятностей вершины u быть в том или ином состоянии. Эти значения меняются во времени по определённым выше правилам, и по ним можно судить о динамике развития эпидемии на графе. Обозначим сразу очевидное свойство $p_S^u + p_I^u + p_R^u = 1$.

Также заметим, что любой неориентированный граф можно разбить на компоненты связности. Понятно, что вершины из разных компонент никак не влияют друг на друга, поэтому далее будем считать, что G — связанный граф.

Теперь можно применить мощный аппарат однородных, конечных марковских цепей к нашей простой модели болезни. Для этого надо разбить цепь на сообщающиеся компоненты и понять какие состояния являются существенными. Т.о. нетрудно прийти с следующим результатам:

1. состояние, в котором все вершины "здоровые" представляет из себя сообщающуюся компоненту, причём единственно существенную в цепи, т.к. из неё недостижимо ни одно другое состояние. Обозначим это состояние как all_S
2. состояния, где все вершины либо S , либо R также образуют индивидуальные компоненты и являются несущественными
3. последняя сообщающаяся компонента состоит из конфигураций, где есть хотя бы одна вершина в состоянии I . Каждое состояние цепи здесь также несущественно

Также легко понять, что все состояния являются апериодическими (всегда есть ненулевая вероятность вернуться в себя). Отсюда, по теореме о солидарности, следует, что все состояния, кроме all_S являются невозвратными и нулевыми, т.е. асимптотически цепь стремиться к тому, что все вершины в графе здоровы с вероятностью 1. Комментарии к этому результату есть в разделе *Обсуждение*.

Динамика заболеваемости

Здесь нам хотелось бы уточнить, как именно меняются со временем распределения вероятностей в вершинах и матожидания количества больных в графе, т.е. как быстро вершины обращаются в здоровые.

Для начала найдём рекуррентные соотношения на распределения **вероятностей в вершинах**. Пусть мы знаем все распределения на предыдущем шаге, тогда для каждой вершины $v \in V$:

$$\mathbb{P}(v \in S_{t+1} \mid \text{знание пред. шага}) = \gamma \mathbb{P}(v \in R_t) + \mathbb{P}(v \in S_t) \cdot \underbrace{\prod_{u \in N(v)} (1 - \mathbb{P}(u \in I_t) \beta w_{uv})}_{A_v}$$

$$\mathbb{P}(v \in I_{t+1} \mid \text{знание пред. шага}) = (1 - \sigma) \mathbb{P}(v \in I_t) + \mathbb{P}(v \in S_t) \cdot (1 - A_v)$$

$$\mathbb{P}(v \in R_{t+1} \mid \text{знание пред. шага}) = \sigma \mathbb{P}(v \in I_t) + (1 - \gamma) \mathbb{P}(v \in R_t)$$

Нетрудно понять, что сумма этих вероятностей равна 1, как и должно быть. Аналогично найдём рекуррентные соотношения для матожиданий:

$$\mathbb{E}(\#S_{t+1}) = \gamma \mathbb{E}(\#R_t) + \sum_{v \in V} \mathbb{P}(v \in S_t) A_v$$

$$\mathbb{E}(\#I_{t+1}) = (1 - \sigma) \mathbb{E}(\#I_t) + \mathbb{E}(\#S_t) - \sum_{v \in V} \mathbb{P}(v \in S_t) A_v$$

$$\mathbb{E}(\#R_{t+1}) = \sigma \mathbb{E}(\#I_t) + (1 - \gamma) \mathbb{E}(\#R_t)$$

Очевидное свойство $\mathbb{E}(\#S_{t+1}) + \mathbb{E}(\#I_{t+1}) = \mathbb{E}(\#R_{t+1}) = \mathbb{E}(\#S_{t+1} + \#I_{t+1} + \#R_{t+1}) = |V|$.

Вычислительный эксперимент

Главная задача вычислительного эксперимента — проиллюстрировать эффект локдауна, смулировав эпидемию на графах контактов, в которых такой эффект вообще возможен. Т.о. данная демонстрация служит подтверждением неголословности оговоренного ранее.

Описание проведения эксперимента

Опишем постановку *экспериментальной части*: создаётся два графа контактов на N вершинах со взвешенными рёбрами. Вес любого ребра $w_{(a,b)} \in [0, 1]$ и интерпретируется как доля времени, проведённая вершиной a с вершиной b за всё время на данном графе контактов. Задаются гиперпараметры эпидемии: вероятности перехода между подверженным/больным/выздоровевшим для вершины. Задаётся начальное распределение больных/здоровых вершин. Далее, вне локдауна, одна итерация для эпидемии проходит так: сначала активен «рабочий» граф, в котором согласно вычисленным вероятностям вершины меняют состояния. Далее становится активен «домашний» граф, на котором

происходит аналогичные действия. При введении карантина же одна итерация будет происходить два раза на «домашнем» графе. По таким правилам эпидемия эволюционирует любое заданное время с сохранением истории состояний для каждой вершины.

Программный пакет SEIRSpluS

Для начала были проведены симулирования заболеваемости с помощью библиотеки SEIRSpluS [7], которая предоставляет богатые средства инициализации модели, а также её тонкой настройки в любой момент развития эпидемии. Стандартный граф контактов генерировался на заданном наборе вершин $V = \{1, \dots, N\}$ и содержал случайное количество рёбер, которое было точно больше половины числа рёбер в полносвязном графе. Граф карантина же собирался из клик случайного размера от 1 до 5 вершин. Чувствительность к заболеванию генерировалась случайно для каждой вершины, со средним вокруг значения 0.5, вероятность восстановления у заболевших вершин была одна для всех 0.3.

В итоге запускалось две симуляции, в одной из которых карантин не вводился, а в другой вводился на некоторое время. Результаты приведены на рис. 1. Здесь представлена динамика количества заражённых узлов для размера популяции 10, 100, 500 вершин в графе. Синяя линия — развитие болезни без локдауна, красная и розовая — развитие с локдауном, где розовая линия как раз отвечает периоду изоляции. Зелёными линиями обозначены точки входа и выхода из локдауна.

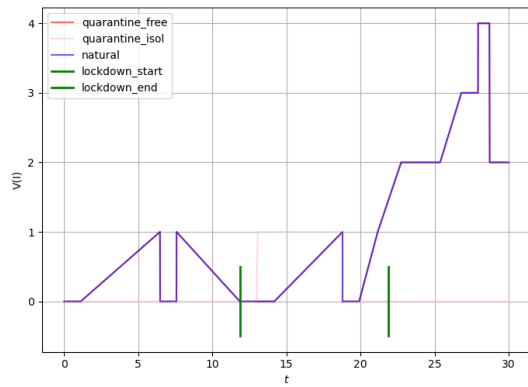
К сожалению, данная библиотека хоть и обладает огромным потенциалом, но всё же является технически недоработанной, а также скрывает в себе некоторые нежелательные методы ускорения сэмплирования, что не является приемлемым для текущего исследования.

Симуляция распространения эпидемии

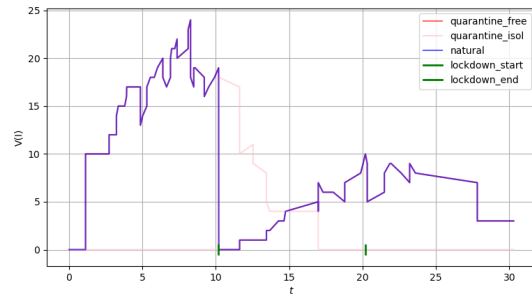
В итоге для данной работы был составлен собственный программный пакет для создания эпидемий и проведения сэмплирования на них. Эпидемия проходит на двух графах контактов, отвечающих за «рабочий» и «домашний» режимы для вершин, во времени, в которой элементарная единица есть один день. Причём если не введён локдаун, то половину дня эпидемия проводится на одном графе, а половину на другом. Вершины меняют состояния согласно вероятностям, заданным по правилам модели. Также задаются временные рамки, в течении которых вводится локдаун, который симулируется как проведение эпидемии только на «домашнем» графе.

Главная цель вычислительного эксперимента — проиллюстрировать эффект локдауна для эпидемий и графов с параметрами, для которых этот самый эффект обязан проявиться согласно теоретическим оценкам.

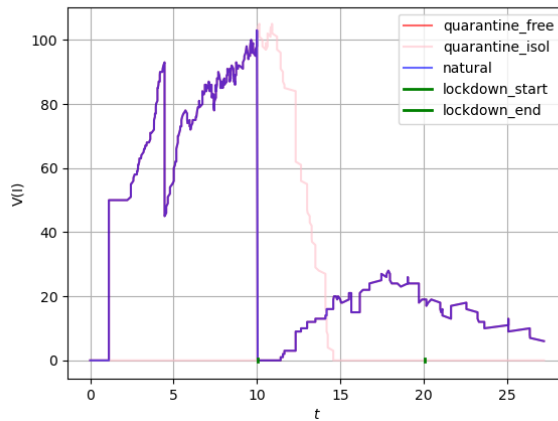
Результаты сэмплирования для нескольких графов представлены на рис. 2 и рис. 3. Основные графы представляли собой полносвязанный и полный цикл на 10 и 15 вершинах, графы карантина представляли собой совокупность клик размера 3 и 4 соответственно. Каждая 3 и каждая 4 вершина в графах были изначально заражены. После проведения многих опытов по описанной выше схеме и усреднения были получены треки эволюции заболеваемости, в которых сравниваются количества заболевших в эпидемии с карантинном и идущей параллельно от неё (до ограничения болезни развиваются одинаково) без карантина. Из данных графиков ясно видно, что для данных конфигураций эпидемия проходит тяжелее во время локдауна и некоторое время после него.



Small population



Medium population



Large population

Рис. 1: Графики количества заражённых в двух режимах протекания эпидемии от времени

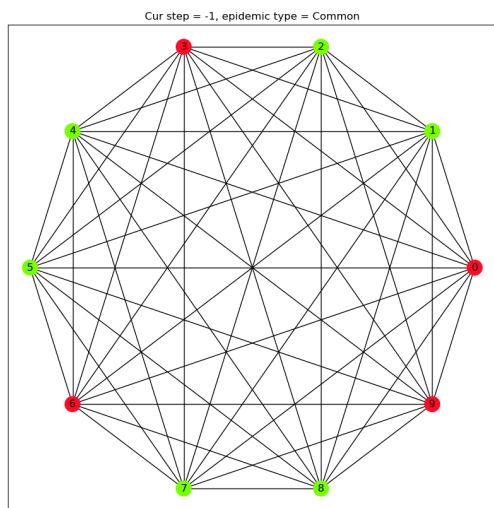
Обсуждение

Полученный сильный результат о том, что цепь сходится к состоянию, где все вершины "здоровы" является побочным эффектом простоты рассматриваемой модели. Достаточно добавить немного реалистичности, например, состояния "гибели" вершин, чтобы асимптотический результат был уже не так просто предсказуем. Действительно, в этом случае состояния цепи, где все вершины здоровы или погибли, также будут существенными и образовывать индивидуальные неразложимые компоненты. Тогда, каково будет распределение вероятностей в пределе между этими состояниями для некоего начального состояния в общем случае трудно предсказуемо.

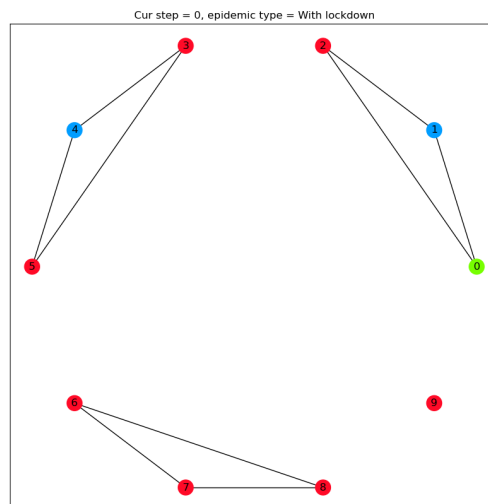
Главный результат данной работы — получения оценок на динамику количества больных в эпидемии на конечных графовых структурах и подтверждение существования эффекта локдауна. Рассматривая формализм марковских цепей, мы выявили, что в простейшей модели эпидемия рано или поздно затухнет. Также мы провели эксперименты на простых графах и выявили, что введение карантина в среднем увеличивает количество больных во время эпидемии и после неё.

Список литературы

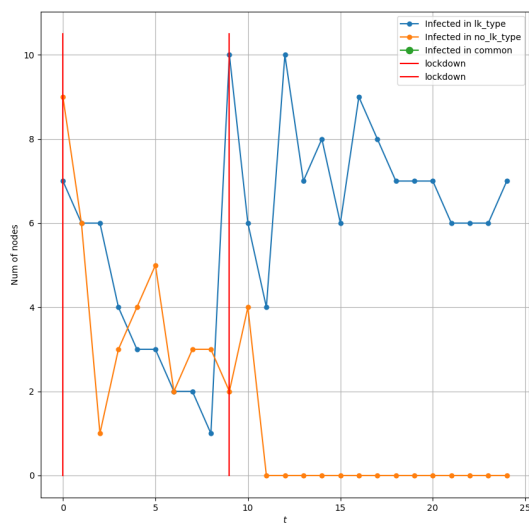
- [1] Linda JS Allen. «Some discrete-time SI, SIR, and SIS epidemic models». В: *Mathematical biosciences* 124.1 (1994), с. 83—105.
- [2] Vincenzo Capasso. *Mathematical structures of epidemic systems*. Т. 97. Springer Science & Business Media, 2008.
- [3] Vittoria Colizza и др. «Epidemic modeling in complex realities». В: *Comptes Rendus Biologies* 330.4 (2007). L'épidémiologie : une science en développement / Epidemiology : a developing science, с. 364—374. ISSN: 1631-0691.
- [4] Manlio De Domenico и др. «The anatomy of a scientific rumor». В: *Scientific reports* 3.1 (2013), с. 1—9.
- [5] Sergio Gómez и др. «Discrete-time Markov chain approach to contact-based disease spreading in complex networks». В: *Europhysics Letters* 89.3 (2010), с. 38009.
- [6] Tiberiu Harko, Francisco SN Lobo и MK3197716 Mak. «Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates». В: *Applied Mathematics and Computation* 236 (2014), с. 184—194.
- [7] <https://github.com/ryansmcgee/seirsplus>, ред. *SEIRS+ library*.
- [8] Jack Leitch, Kathleen A Alexander и Srijan Sengupta. «Toward epidemic thresholds on temporal networks: a review and open questions». В: *Applied Network Science* 4 (2019), с. 1—21.
- [9] Yamir Moreno, Romualdo Pastor-Satorras и Alessandro Vespignani. «Epidemic outbreaks in complex heterogeneous networks». В: *The European Physical Journal B-Condensed Matter and Complex Systems* 26 (2002), с. 521—529.
- [10] Romualdo Pastor-Satorras и др. «Epidemic processes in complex networks». В: *Reviews of modern physics* 87.3 (2015), с. 925.
- [11] Бишук. «Применение активного обучения к графовым моделям на примере оценки рисков распространения эпидемии». Диплом бакалавра. МФТИ, 2021.



Начальная конфигурация вершин и вид
«рабочего» графа

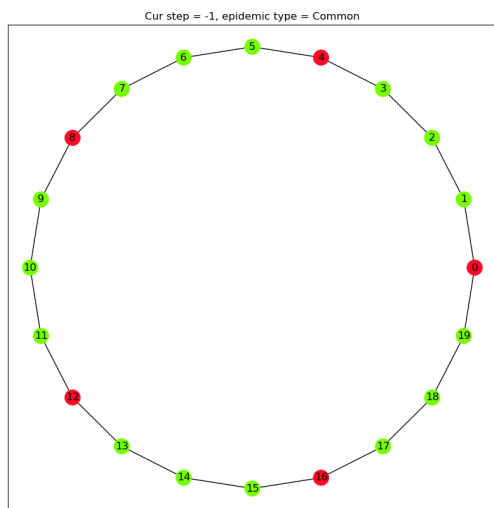


Вид «домашнего» графа

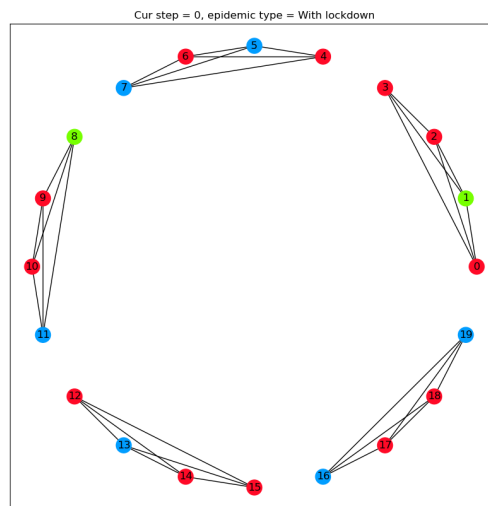


Сравнительные треки эпидемии с
локдауном и без него

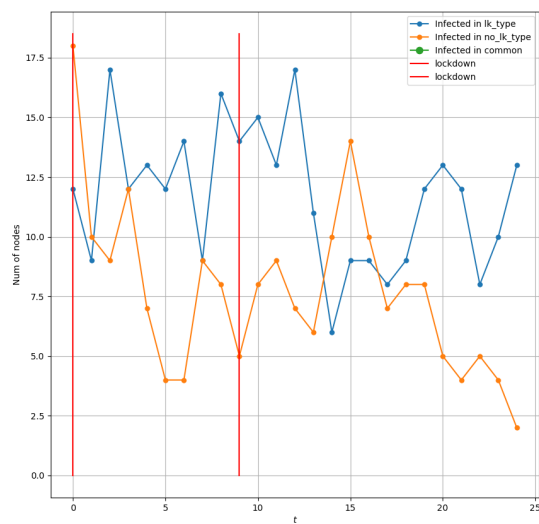
Рис. 2: Визуализация протекания эпидемии для полносвязанного графа



Начальная конфигурация вершин и вид
«рабочего» графа



Вид «домашнего» графа



Сравнительные треки эпидемии с
локдауном и без него

Рис. 3: Визуализация протекания эпидемии для графа-цикла