АННОТАЦИЯ

Задержки в работе транспортных сетей имеют негативные последствия для менеджеров инфраструктуры и служб обслуживания, а также для путешественников. Хотя широко признано, что задержки, возникающие в транспортных сетях, могут быть связаны, существует недостаток знаний о базовых свойствах этих взаимосвязей и средствах их количественной оценки. С этой целью мы разрабатываем общесетевой метод анализа задержек на основе данных. Сначала мы строим байесовскую сеть для представления соотношений между задержками, связанными с различными элементами транспортной сети, и оценки надежности критически важных элементов инфраструктуры. Во-вторых, мы предлагаем серию оригинальных метрик, называемых показателями информативности, для количественной оценки пространственной протяженности наблюдаемых задержек на основе полученной байесовской сети. Предлагаемый подход применяется к сети метро Вашингтона, округ Колумбия. Зависящие от времени задержки ожидания пассажиров и пересадки, рассчитанные более чем на год на основе данных смарт-карты, используются в качестве входных данных для байесовской сети. Наши результаты показывают, что задержки пассажиров на нескольких выбранных станциях напрямую информируют о задержках, происходящих на многих других станциях. Мы также исследуем связь между предлагаемыми показателями информативности и топологическими свойствами станций метро, приходя к выводу, что последние имеют ограниченное значение при аппроксимации общесетевых корреляций задержек.

1. Введение

Хотя задержки в транспортных сетях наблюдаются на местном уровне, они имеют последствия для общей надежности сети. Менеджеры инфраструктуры и поставщики услуг заинтересованы в минимизации задержек и разработке эффективных мер по их снижению. Для эффективного повышения надежности транспортировки важно понимать, как взаимосвязаны задержки по всей сети. Это особенно важно в сетях общественного транспорта, где задержки могут передаваться каскадом не только на соседние сегменты, но и далее по потоку от вышестоящих служб, а также между службами из-за взаимозависимости активов и подвижного состава, а также схемы обмена пассажирами. В прошлых работах по распространению задержек к этой проблеме подходили с точки зрения оператора сети, а не пассажира [2, 17]. Маниц и др. [26] предложили метод определения источника задержек поездов, основанный на анализе эпидемиологических процессов. Кафле и Цзоу [16] предложили аналитическую модель для моделирования распространения задержек рейсов в зависимости от расписания полетов данного самолета. Анализ временных рядов был проведен Du et al. [7] для анализа распространения задержек на уровне аэропорта. Харрод и др. [13] сформулировали аналитическую модель для оценки распространения задержек транспортных средств вдоль железнодорожной линии. Даниэле Марра и Корман [27] определили кластеры задержек в данных автоматического определения местоположения транспортного средства (AVL), используя вариант алгоритма DBSCAN. Бюхель и др. использовали анализ временных рядов и стохастическую имитационную модель железнодорожных операций. [3] для анализа распространения задержек в сети железных дорог Швейцарии. Подводя итог, предыдущие исследования были сосредоточены на анализе пунктуальности транспортных средств и ее распространении по отдельным маршрутам, при поездках данного транспортного средства или по сети. В этих исследованиях использовался набор аналитических, статистических методов и методов моделирования для анализа задержек в системах пассажирского транспорта, однако мало что известно о свойствах задержек в масштабах всей сети, с которыми сталкиваются пассажиры.

Пассажирские перевозки состоят из нескольких последовательных сегментов. Время в пути, а следовательно, и задержки, складываются из времени ожидания, нахождения в транспортном средстве и, возможно, пересадки. Таким образом, расстояние между транспортными средствами, ограниченная вместимость транспортных средств и синхронизация передач являются одними из факторов, определяющих задержки пассажиров. В литературе по оценке надежности сети общественного транспорта часто используются обобщенные потери времени в пути пассажиров в качестве показателя уязвимости сети и средства идентификации наиболее важных звеньев в сети [30, 4]. Модели динамического назначения позволяют оценить пространственную степень распространения задержек в сетях общественного транспорта путем анализа сценариев сбоев [25]. Весь вышеупомянутый анализ надежности был выполнен либо на основе аналитической, либо на основе имитационной модели назначения общественного транспорта. Более того, сценарный анализ основывался на полных сбоях, в то время как задержки пассажиров были вызваны множеством мелких нарушений. Кроме того, взаимосвязь между задержками, происходящими в разных местах, не изучалась. Таким образом, остается неизвестным, как на самом деле связаны задержки пассажиров в разных точках сети. Возможно, это связано с отсутствием эмпирических знаний о траекториях движения пассажиров, а также с необходимостью разработки методов и приемов для определения задержек пассажиров и систематической количественной оценки их взаимосвязей. Анализ таких соотношений позволит выявить пространственные закономерности задержек, происходящих одновременно по всей сети, и внести вклад в анализ надежности системы. Измерение задержек пассажиров становится все более возможным посредством обработки пассивно собранных данных. В нескольких исследованиях были проанализированы задержки в поездках пассажиров с использованием записей смарт-карт [34, 14]. Несмотря на это, существует недостаток эмпирических знаний о свойствах задержек пассажиров в масштабах всей сети и, в частности, о взаимосвязи между задержками, происходящими по всей сети. Для этого необходимы данные о задержке пассажиров, связанные с отдельными сегментами поездки, т.е. элементами сети. Кришнакумари и др. [20] разработали метод оценки для определения задержки, связанной с каждым компонентом поездки, т. Е. Временем ожидания, временем нахождения в транспортном средстве, временем передачи на основе объединения пассивно собранных данных, касающихся как движения транспортного средства, так и пассажиров (и без использования каких-либо файлов журнала сбоев). Это открывает возможности для изучения взаимосвязей между этими задержками и выявления потенциальных закономерностей, которые могут быть использованы разработчиками моделей и планировщиками. Основываясь на задержках пассажиров, выведенных из данных смарт-карт, Yap и Cats [33] применяют контролируемый обучающий подход для прогнозирования задержек пассажиров на отдельных станциях и периодах времени и определяют наиболее критичные станции с точки зрения подверженности, воздействия и общей критичности (измеряемой в потерях времени прохождения). Затем результаты были использованы для кластеризации остановок на основе их вклада в общую уязвимость сети, а анализ характеристик выявил наиболее важные переменные для прогнозирования воздействия сбоев.

Байесовские сети ранее использовались для оценки риска и надежности критически важной инфраструктуры в различных областях применения, от захоронения ядерных отходов [22] и химических мощностей в условиях преднамеренных атак [1] до анализа морских аварий [9], дорожных сетей [11] или систем водоснабжения и электроснабжения [21] в условиях сейсмической опасности, а также портов внутреннего водного транспорта и прилегающей сети поставок [15]. В контексте железнодорожных операций байесовские сети использовались Зилко и др. [35] и Лессаном и др. [23] для прогнозирования продолжительности сбоев и задержек поездов соответственно. Гаэми и др. [12] продемонстрировало, как результаты байесовского сетевого прогнозирования могут быть интегрированы в модель управления сбоями для прогнозирования последствий сбоев, а также мер по изменению расписания в случае задержек пассажиров. Улак и др. [32] предложили использовать две сетевые метрики для различения станций, вызывающих задержки, и станций, подверженных задержкам, и применили их для анализа задержек пригородных поездов. Следуя прошлым исследованиям, мы применяем подход, основанный на данных, который использует принципы машинного обучения и науки о сетях. Этот подход позволяет устанавливать статистические взаимосвязи между элементами системы, одновременно устраняя необходимость делать множество предположений о лежащих в основе взаимосвязях между компонентами системы. Это позволяет выявлять неожиданные взаимосвязи и учитывать все возможные зависимости, задействованные в механизме распределения задержек.

Ни в одном из вышеупомянутых исследований не анализировались общесетевые закономерности задержек пассажиров и их пространственное соотношение. Насколько нам известно, это первое исследование, в котором эмпирически изучаются эти общесетевые взаимосвязи между задержками пассажиров. В этом исследовании мы разрабатываем метод анализа общесетевых задержек пассажиров и их каскадных моделей. Результаты этого исследования заключаются в следующем: (i) оценить байесовскую сеть задержек пассажиров, связанных с отдельными элементами сети; (ii) разработать серию оригинальных показателей, называемых индикаторами информативности, для количественной оценки пространственных масштабов наблюдаемых задержек на основе полученных топологических характеристик байесовской сети; (iii) применить предлагаемый метод к обширным эмпирическим данным о задержках пассажиров из сети метро Вашингтона, округ Колумбия; (iv) изучить связь между предлагаемыми показателями информативности и топологическими свойствами станций метро.

Предлагаемые показатели информативности позволяют количественно оценить способность станций предоставлять информацию о состоянии задержки в остальной части сети. Это позволяет нам установить степень, в которой данное состояние станции с точки зрения количества наблюдаемых там задержек пассажиров может указывать на задержки, наблюдаемые в других частях сети. Мы демонстрируем, как эти индикаторы содержат ценную информацию, которая потенциально может быть использована для определения критических мест для применения мер по снижению задержек, а также для предоставления информации как операторам, так и пассажирам. Чтобы пролить свет на основные процессы, связанные с обнаруженными корреляциями задержек пассажиров, мы анализируем связь между рассчитанными показателями информативности и рядом топологических показателей физических сетей и сетей обслуживания, которые отражают различные функции системы. Это также позволяет определить, могут ли топологические индикаторы потенциально использоваться в качестве прокси для предлагаемых показателей информативности в отсутствие подходящих данных.

Остальная часть этой статьи структурирована следующим образом: метод, включая предлагаемые показатели информативности, описан в разделе 2; затем применение этого исследования, включая описание тематического исследования, кратко представлено в разделе 3; Результаты представлены и обсуждены в разделе 4, а выводы сделаны в разделе 5.

2. Метод

Входные данные для нашего метода включают расчетную задержку пассажиров для каждого элемента сети (станции, сегмента пути) и несколько графических представлений соответствующей сети. Сначала мы определяем взаимосвязи между задержками, возникающими в разных местах, посредством построения байесовской сети (раздел 2.1). После этого мы определяем количественно степень совпадения задержек, возникающих в одном элементе сети, с задержками, возникающими в других местах, путем вычисления ряда исходных показателей, которые включают информацию, относящуюся к топологии полученной байесовской сети, в то, что мы называем информационными показателями (раздел 2.2). Наконец, мы перепроверяем связь между полученными информационными показателями и топологическими показателями самой сети общественного транспорта (раздел 2.3).

2.1. Определение соотношений задержек: создание байесовской сети

Сначала на основе доступных данных строится байесовская сеть (BN). A BN по определению является направленным ациклическим графом (DAG) - G (N, A), где N - набор узлов, а A - набор дуг, который представляет взаимозависимости его переменных. Здесь узел, у которого есть дуга, указывающая на другой узел, называется родительским, а узел, на который направлена дуга, называется дочерним. A BN также содержит информацию о зависимости посредством таблиц условной вероятности состояния переменной, где вероятность возникновения состояния зависит от наблюдаемого состояния соответствующих родительских элементов. Для получения дополнительной информации о BN заинтересованный читатель может обратиться к Кьерульфу и Мадсену [18] и Коллеру и др. [19].

В этом исследовании набор потенциальных узлов в BN соответствует набору станций, где каждое направление линии обозначено как отдельный узел станции. Дуги графика представляют взаимосвязи между узлами независимо от физических или сервисных соединений между станциями в транспортной сети. Среднее время ожидания пассажира на каждой станции используется в качестве входных данных и получается из алгоритма оценки, применяемого к индивидуальным траекториям движения пассажиров, как подробно описано в описании подготовки входных данных в разделе 3.1.

Используя таблицы условной вероятности, дуги BN могут быть помечены в соответствии с силой корреляции между каждой парой соединенных узлов. Маркировка выполняется с использованием метода силы связи, представленного Николсоном и Джитнахом [28]. В окончательном представлении BN сохраняются только дуги с метками, отражающими взаимозависимость, измеряемую с помощью коэффициента корреляции, которая превышает определенную значимость. Наиболее распространенный подход к маркировке дуг BN основан на взаимной информации (MI)) [29, 28, 18, 8], который количественно определяет, какой объем информации может быть получен относительно определенного узла, как только информация, касающаяся другого узла, становится доступной. Эта концепция может быть расширена, чтобы включить возможность наличия у узла нескольких родительских узлов, так что при определении силы дуги учитываются взаимозависимости между родительскими узлами [8, 28]. Метка силы связи rij для связи, соединяющей узлы i и j, оба элемента набора узлов V, может быть рассчитана следующим образом:



Где x и y - векторы, представляющие все возможные состояния узлов i и j соответственно. Аналогично, z является единственным вектором, содержащим все возможные состояния всех родительских узлов узла j, за исключением i. rij передает, таким образом, силу зависимости между двумя узлами в графе BN, независимо от значения других родительских узлов нисходящего узла. Термины вероятности в уравнении. (1) относится к состояниям переменных, например, p (y | x, z) - это (условная) вероятность нахождения значений, связанных с узлом j, в любом конкретном состоянии при определенной комбинации состояний переменных в узле i и всех других родительских узлах.

2.2. Показатели информативности

Далее мы используем структуру меток BN и arc для расчета показателей информативности для каждого узла. Далее мы представляем набор исходных показателей. Предлагаются разные показатели, поскольку они охватывают разные понятия информативности и могут привести к разным наблюдениям. Они определены и кратко обсуждены ниже.

Первый предлагаемый показатель информативности - это степень исходящего узла. В данном контексте этот показатель предоставляет информацию о том, на сколько узлов определенный узел может предоставлять прямую информацию, и рассчитывается следующим образом



где d +i - исходящая степень узла узла i ∈ V; N - набор узлов в BN; и ai,j ∈ A - вход в матрицу смежности A графа BN, который равен 1, если есть ссылка, соединяющая узел n и узел i, и 0 в противном случае.

Хотя степень исходящего узла указывает, о скольких узлах определенный узел предоставляет информацию, она не отражает объем этой информации. Поэтому мы предлагаем в качестве второго показателя ожидаемую прямую информативность. Оно предназначено для описания того, насколько информативным, как ожидается, будет определенный узел в отношении любого узла, к которому он подключен. Ожидаемая прямая информативность, ei, рассчитывается с использованием средневзвешенного показателя:



Два вышеупомянутых индикатора дают представление только о количестве информации, которая может быть предоставлена через непосредственных нижестоящих соседей (также известных как ‘дочерние элементы’ в терминологии BN) узла. Однако узел также может косвенно предоставлять информацию о своих дальнейших потомках через отношения более высокого порядка. Это можно учесть путем умножения весов ссылок вдоль пути, соединяющего его с потомком. Когда существует несколько возможных путей, мы рассматриваем два

варианты: (i) может быть принят во внимание наиболее информативный путь или (ii) могут быть приняты во внимание все пути, поскольку разные пути могут предоставлять новую информацию. В первом случае информативность может быть недооценена, поскольку учитываются не все информационные пути. И наоборот, в последнем случае показатель информативности может быть завышен, поскольку некоторые пути могут перекрываться, т.е. Не являются независимыми, коррелированными. Следовательно, мы применяем оба подхода и используем их для установления нижней и верхней границ информативности данного узла для всех его потомков.

Чтобы определить общую информативность, то есть общую информацию, которую узел может предоставить о состоянии задержки остальной части сети, значение информативности для всех потомков узла в сети должно быть суммировано.

Мы позволяем (k, m) ∈ p означать, что существует связь с вышестоящим узлом k и нижестоящим узлом m (т.е. ak,m = 1; k, m ∈ V), которая принадлежит пути p. Верхняя граница этого показателя, tmaxi , может быть выражена как



где Pi,j - набор всех путей, соединяющих узлы i и j.

Нижняя граница этого показателя, tmini, формулируется как



2.3. Связь показателей центральности с показателями информативности

Построение BN, на основе которого рассчитываются показатели информативности, требует большого объема эмпирических данных, которые часто недоступны. Более того, мы заинтересованы в исследовании связи между найденными коэффициентами задержки пассажиров и функциональными характеристиками системы. Таким образом, третий этап методологии направлен на тестирование потенциального использования топологических индикаторов, которые не требуют никаких данных, кроме статических конфигураций сети и сервисов, в качестве прокси для сбора информации об узлах. Луо и др. [24] предоставляют доказательства того, что строго топологические индикаторы, т. е. индикаторы центральности станций, позволяют прогнозировать распределение пассажиропотоков по сети, особенно с учетом уровня обслуживания, включая количество пересадок и частоту обслуживания, подразумеваемые конструкцией сети обслуживания, в дополнение к уровню инфраструктуры.

Мы вычисляем корреляции трех предложенных показателей информативности, используя серию показателей центральности узла, т. е. степень централизации узла, близость и промежуточность для трех различных вариантов графика сети общественного транспорта на основе того, что иногда называют в литературе L-, P- и B-пространствами [6, 10]. Настоящим мы обозначаем их как Инфраструктурные, Сервисные и Транспортные пространства соответственно. Эти пространства представляют физический и логический уровни системы общественного транспорта. Пространство инфраструктуры представляет все станции как узлы, и любая пара станций, которые последовательно обслуживаются данной линией, соединены ссылкой. Хотя в Сервисном пространстве используется один и тот же набор узлов, связи вводятся между всеми парами остановок, которые обслуживаются по крайней мере одной общей линией, независимо от их положения вдоль линии. Пространство переноса содержит два набора узлов: в дополнение к узлам станции каждая строка представлена узлом. Ссылки вводятся только между узлами, принадлежащими к разным подмножествам – между станцией и линией - обозначая, что станция обслуживается данной линией.

Различные графические представления позволяют идентифицировать характеристики станций с высоким значением информативности. Один и тот же показатель центральности узла имеет разные значения при применении в разных пространствах, все из которых могут иметь отношение к информативности. Например, центральность по степени узла соответствует количеству прямых соседей на графике. В зависимости от рассматриваемого пространства это значение может интерпретироваться по-разному: в инфраструктурном пространстве степень узла отражает локальную роль пересечения / взаимозаменяемости; в служебном пространстве это соответствует количеству других станций, на которые влияет та же линия, и; в пространстве передачи это количество линий, пересекающихся на данной станции (или количество станций, обслуживаемых линией, в случае линейного узла). Аналогично, центральность между узлами соответствует доле кратчайших путей, проходящих через узел. В то время как в Инфраструктурном пространстве он может использоваться для определения пассажиропотоков, проходящих через станцию, в Сервисном пространстве он отражает вероятность того, что станция использовалась в качестве пересадочного узла.



Рисунок 1- Карта сети метро Вашингтона, округ Колумбия, с нашим кодированием всех станций и указаниями маршрутов пересадок.

3. Заявка

Метод, описанный в предыдущем разделе, применяется к тематическому исследованию сети общественного транспорта, для которой доступны индивидуальные маршруты движения пассажиров и задержки пассажиров (раздел 3.1). После представления характеристик задержки пассажиров (раздел 3.2) мы обсуждаем результаты трех ключевых методологических этапов, а именно результаты BN (раздел 3.3), значения показателей информативности (раздел 3.4) и их корреляции с топологическими показателями для сети тематических исследований (раздел 3.5).

3.1. Тематическое исследование и реализация в метро Вашингтона, округ Колумбия

Для этого исследования доступны выходные данные о движении поездов за год и назначении пассажирских поездов с помощью так называемого метода ODX, описанного в Sanchez-Martinez [31], из системы метро Вашингтона, округ Колумбия (рис. 1), управляемой Управлением городского транспорта Вашингтона, округ Колумбия (WMATA). Данные состоят из данных смарт-карты за один год с 19 августа 2017 года по 28 августа 2018 года для всей системы метро, которая состоит из шести линий, обслуживающих 91 станцию.

WMATA определяет для каждой пары станций отправления и назначения в их сети метро максимальное ожидаемое время в пути. Последнее соответствует случаю, когда пассажир только что опоздал на поезд, при условии, что следующий поезд прибывает в пределах запланированного времени обслуживания, время, проведенное на борту, соответствует запланированному времени, а время ходьбы соответствует 80-му процентилю распределения скорости ходьбы. Чтобы определить это математически, давайте обозначим через ts0, sd, τ максимальное ожидаемое время в пути для поездки rso, sd которое происходит в течение периода времени τ, где so ∈ S и sd ∈ S - это исходная (вводная) и конечная (выводная) станции этого маршрута соответственно. S - набор всех станций. Максимальное ожидаемое время в пути состоит из следующих трех компонентов:

Первое слагаемое суммирует все запланированное время в пути в транспортном средстве, tвместоsi, si + 1, τ, между последовательными станциями (т. е. si, si + 1 ∈ S), пройденными линией l, которая пересажена на соответствующем маршруте. Второй термин - это планируемая скорость обслуживания каждой из линий, загруженных на соответствующем маршруте. Третий термин соответствует всем пешеходным сегментам на протяжении всего пути в системе метро (т. е. между местом входа и посадочной платформой, между платформами выхода и посадки в случае пересадок и между платформой выхода и местом выхода).

Общая задержка в пути пассажира рассчитывается как время в пути, превышающее максимальное ожидаемое время в пути для соответствующей пары станций ввода и вывода. Задержка, связанная с поездкой пассажира n, которая включает в себя поездку между станцией отправления s0 и станцией назначения sd, которая имела место в течение периода времени τ, раскладывается следующим образом:



где dwait so , τ − задержка, связанная с ожиданием на станции отправления, dбортовойsi, si+1, τ - задержка, возникающая при прохождении определенного сегмента поезда, d transsi, τ - задержка, связанная с местом пересадки, а en - оставшаяся задержка для конкретной поездки, которая не связана ни с одним из элементов сети. Кришнакумари и др. [20] сформулировали набор уравнений для сопоставления задержек в поездках пассажиров с задержками пассажиров, связанными с каждым элементом сети. Их выходные данные доступны в качестве входных данных для этого исследования.

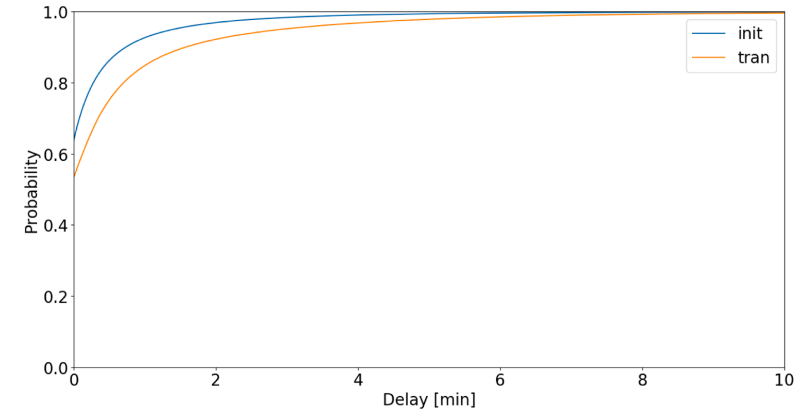


Рисунок 2- Функция кумулятивного распределения начальной задержки и задержки пересадочного пассажира.

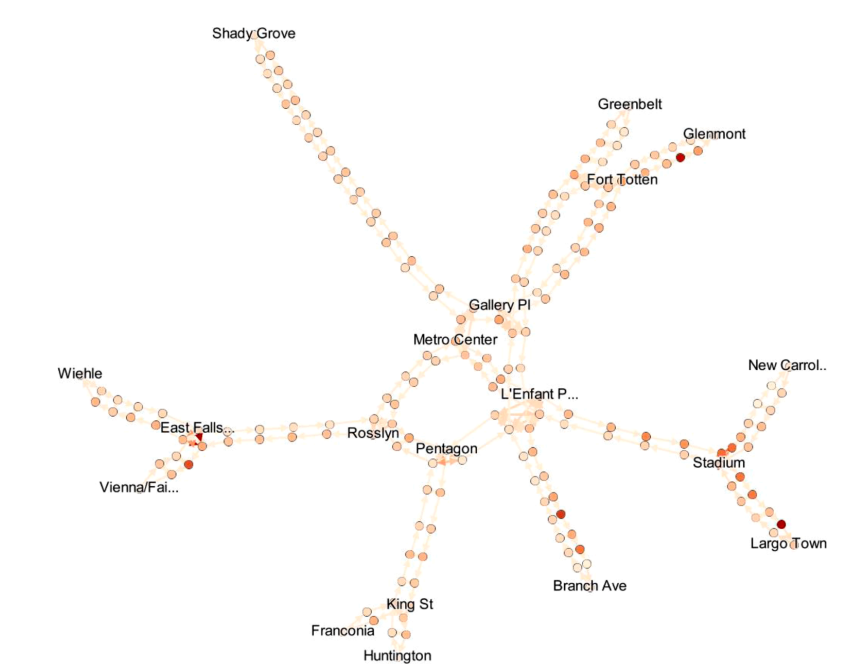


Рисунок 3 - Средняя задержка пассажиров на каждой станции отправления, пересадочной станции и пересадочном узле. Более темный цвет указывает на более длительные задержки.

Таким образом, исходные данные, доступные для этого исследования, состоят из расчетной средней начальной задержки пассажиров и задержки пересадки на каждой станции и направлении пересадки, в результате чего в нашем графическом представлении получается 186 узлов. Обратите внимание, что компоненты задержки пассажиров, связанные с пересадочной станцией, которая позволяет осуществлять пересадку между l линиями, представлены с использованием 2l узлов начального времени ожидания (по одному на направление линии) и 2l (2l − 1) − 2l или 4l (l − 1) узлов времени пересадки (по одному на комбинацию направлений линии, исключая противоположные направления одной линии).

Средние задержки пассажиров оценивались за 30-минутный временной интервал в течение всего периода анализа. Данные о задержках пассажиров дифференцированы в зависимости от направления, в котором путешествуют пассажиры. Ниже были сохранены только обычные рабочие дни (т. е. исключая выходные, праздничные дни и будни с масштабными работами по техническому обслуживанию). Это было сделано, поскольку ожидается, что из-за различных характеристик доступности спроса и предложения в исключенные дни соотношение задержек может отличаться и, таким образом, может привести к искажению результатов. Параметру p-value было присвоено стандартное значение 0.05 при принятии решения о том, следует ли включать значимость зависимости в BN.

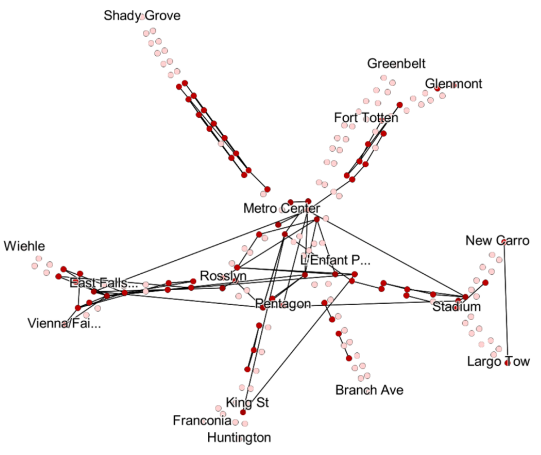


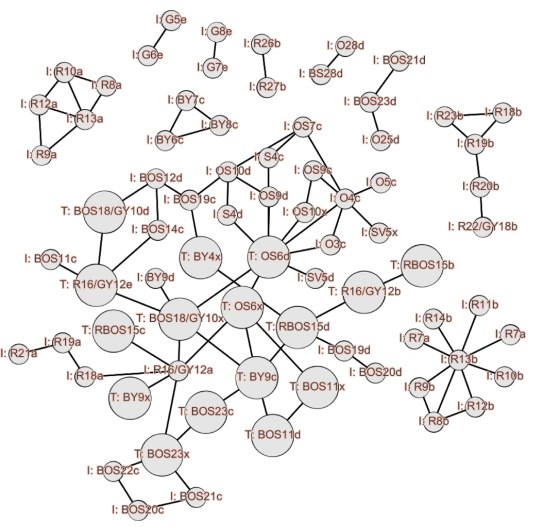
Рисунок 4 – Рекомбинированные BN отображены на географической карте станций. В презентации не проводится различия между передаточными и исходными станциями. Узлы, которые не подключены к BN, показаны светло-розовым цветом, в то время как узлы, которые подключены, показаны темно-красным.

Рисунок 5 - Во всех подключенных узлах байесовской сети узлы передачи отображаются как более крупные, чем узлы начальной станции. Код станции (см. рис. 1) определяется указанием того, относится ли узел к начальному времени ожидания (I:) или ко времени передачи (T:).

BN был создан с использованием библиотеки libpgm для Python 2.7 [5]. Сборка BN заняла 15,5 часов на персональном ноутбуке HP EliteBook 840 G1 с процессором Intel (R) Core i7–4510U с частотой 2,00 ГГц и 2,60 ГГц с оперативной памятью 8 ГБ.

3.2. Описательная статистика задержек пассажиров

Задержки пассажиров сильно различаются по временным срезам и элементам сети. Как видно на рис. 2, распределение задержек таково, что большинству станций не приписаны задержки пассажиров в течение большинства временных срезов. Напоминаем, что определение задержки пассажира (пример (6)) таково, что задержка регистрируется только в том случае, если пассажир ожидает дольше, чем предусмотрено полное время обслуживания. Более того, подавляющее большинство станций, которые действительно вызывают задержки – либо начальные, либо при ожидании – составляют до 2 минут на пассажира. Несмотря на это, за год сбора данных есть несколько случаев, когда отдельные узлы были связаны с задержкой в 5 и даже 10 минут на пассажира.

Далее мы переходим к изучению пространственного распределения задержек пассажиров между узлами сети. На рис. 3 отображается среднее значение (за все 30-минутные временные отрезки) задержки на одного пассажира для каждой начальной станции (представленной узлом для каждого направления), связи (дуги между узлами разных станций) и пересадочного движения (дуга, соединяющая узлы одной и той же станции с разными прямыми маршрутами). Задержки, как правило, равномерно распределяются по сети после усреднения за все периоды времени. Несмотря на это, на некоторых станциях наблюдаются периодические и серьезные задержки с задержкой пассажиров в среднем на несколько (до 4,6) минут при усреднении за весь анализируемый период. Высокие значения наблюдаются на станциях, расположенных в непосредственной близости от конечных станций (т. е. Выше по течению от центра Гленмонта и Ларго), а также при некоторых пересадках (например, в Ист-Фоллс-Черч).

Как видно также из рис. 2, средние задержки при перемещении поездов увеличиваются по сравнению с начальным временем ожидания на станциях. Однако только 36% всех пассажирских поездок связаны с пересадкой, а среднее количество этапов поездки в вашингтонском метро составляет 1,43. Несмотря на то, что станции внутри внутренних колец (Gallery Palace-Metro Center-Rosslyn-Pentagon-L'Enfant Plaza) подвержены задержкам от низких до умеренных в среднем, из-за очень большого пассажиропотока на этих станциях, на них, тем не менее, приходится большая часть задержек пассажиров, происходящих в сети.

3.3. Результаты по байесовским сетям

В оставшейся части этого исследования мы сосредоточимся на задержках, возникающих во время ожидания – либо на начальной станции, либо при пересадке. Каждый элемент в векторе, представляющий все возможные состояния для всех узлов (x в уравнении (1)), соответствует значению задержки для 30-минутного временного среза в течение периода анализа. Из-за размера сети определение байесовской сети для всех станций одновременно является непомерно трудоемким с точки зрения вычислений. Таким образом, сеть разделена на сектора: один сектор для ядра сети и шесть секторов для различных участков, отходящих от ядра сети (например, к западу от Росслина, к югу от Пентагона, см. рис. 1). Для каждого из этих секторов создается BN. При наличии некоторого перекрытия между секторами (а именно, первых двух станций каждого радиального сектора, ближайшего к ядру, а также передаточных станций, которые присутствуют в радиальных секторах), возможно впоследствии рекомбинировать BN секторов в один большой BN. Это упрощение представления BN сделано, поскольку ожидается, что между станциями, расположенными дальше друг от друга, не будет обнаружено значительных зависимостей; гораздо более вероятно, что удаленные друг от друга станции будут зависеть друг от друга только косвенно, через станции, расположенные между ними.

Результаты построения байесовской сети, отображенные с использованием географической карты сети метро, можно увидеть на рис. 4. Из 186 узлов, включенных в график сети метро Вашингтона, 75 соединены в полученном BN с использованием 85 дуг (т. Е. p-значение для соответствующего rij, как определено в уравнении. (1) меньше 0,05).

Можно сделать несколько наблюдений, а именно, что соединения существуют преимущественно между близлежащими станциями и между узлами, которые движутся в одном направлении. Это подтверждает предположение, лежащее в основе построения BN для секторов. Несмотря на это, есть также несколько соединений, которые связывают станции, находящиеся далеко друг от друга. Кроме того, результаты для различных сгибов набора данных (данные были разделены на несколько обучающих и тестовых наборов для k-кратного подхода к вычислению ошибок модели) показали, что средний процент ошибок для всех узлов был значительно ниже 10% при сравнении для различных разбиений набора данных с наибольшей ошибкой в 7,4%, наблюдаемой для пересадочной станции Пентагона в направлении Арлингтонского кладбища. Мы также исследуем полученные результаты для разных складок, и во всех случаях делаются аналогичные наблюдения относительно типов обнаруженных зависимостей, а именно, что большинство зависимостей возникает между соседними станциями и узлами, соответствующими одному и тому же направлению обслуживания. Кроме того, мы вычисляем среднеквадратичную ошибку (RMSE) полученного BN, используя 5-кратный метод. В результате набора тестов среднее значение RMSE составило 0,0092 минуты, без существенных различий между сгибами. Значение RMSE очень низкое, учитывая, что задержки могут составлять от 0 до более чем 5 минут.

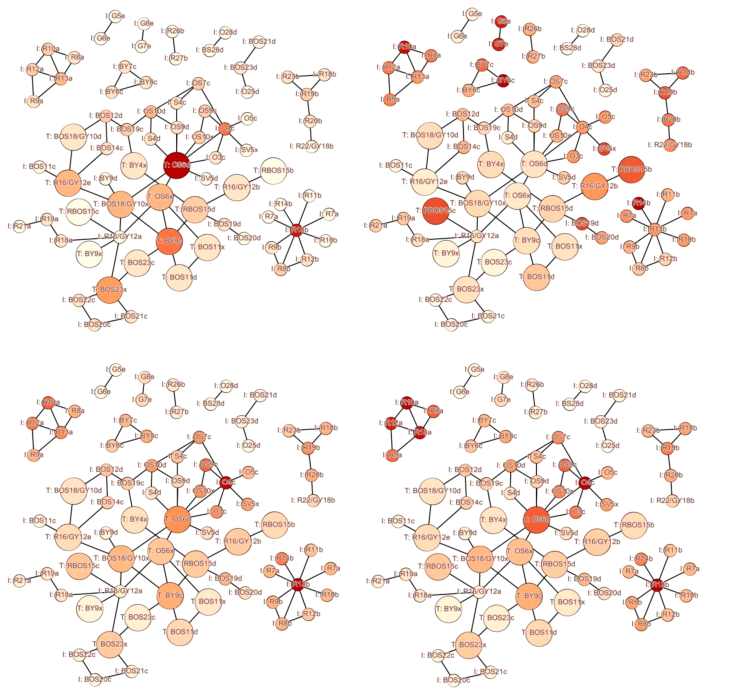


Рисунок 6- Все подключенные узлы в байесовской сети, узлы передачи изображаются как более крупные, чем узлы начальной станции. Цветовая шкала отражает относительное значение соответствующего показателя информативности, показанного по часовой стрелке, начиная с верхнего левого угла, следующим образом: Степень исходящего узла, ожидаемая прямая информативность, Нижняя граница общей информативности и Верхняя граница общей информативности.

Чтобы лучше понять свойства полученного BN, мы отобразим его конфигурацию на рис. 5. Как хорошо видно на рис. 5, полученный BN представляет собой несвязный граф, состоящий из одного большого подкомпонента (42 узла), одного значительно меньшего компонента (9 узлов) и восьми дополнительных компонентов размером от 2 до 5 узлов.

Крупный компонент хорошо виден на рис. 5 состоит как из начального времени ожидания, так и из узлов передачи, каждый из которых имеет более высокую вероятность соединения с другим узлом того же типа (начальный-initial/ transfer-transfer), чем с другим типом (начальный-transfer). Более того, все остальные компоненты состоят исключительно из узлов начального ожидания. Более тщательный анализ также показывает, что в случае восьми самых маленьких компонентов, за одним исключением, все они состоят из набора станций, расположенных в одном направлении линии обслуживания. Исключение относится к паре из двух восточных конечных станций (видны также на рис. 4) Оранжевой (т. е. Нью-Кэрроллтон) и Синей / Серебряной линий (т.е. Центр города Ларго), большая часть линий которых проходит по общему коридору, и поэтому режим их диспетчеризации, вероятно, будет взаимозависимым.

Линии метро имеют тенденцию образовывать клики на графике BN, что указывает на то, что задержки во времени ожидания в основном распределяются между станциями, обслуживаемыми одной и той же линией. Все узлы компонента среднего размера в BN (рис. 4) расположены вдоль Красной линии в любом направлении его северо-западной части. Даже в большом компоненте, хотя он состоит из множества строк, строки имеют тенденцию образовывать клики. Корреляции задержек пассажиров на станциях одной и той же службы, которые не расположены в тесной последовательности, могут быть отнесены к переходам на линии. Обратите внимание, что определенные задержки передачи, подключенные в BN, также являются последовательными станциями вдоль линии обслуживания (например, R16 / GY12b и RBOS15b).

В отличие от задержек транспортных средств, задержки пассажиров вполне могут быть коррелированы также между линиями из-за перемещающихся потоков, объектов, используемых несколькими линиями (т. Е. Терминалов, путей, станций) и системных особенностей, которые выходят за рамки одной линии (например, общий уровень спроса, погодные условия). Действительно, хотя большинство задержек пассажиров связаны на уровне линии, они ни в коем случае не ограничиваются одной линией, как это хорошо видно также на рис. 4.

Интересно, что некоторые отношения не проявляют транзитивности. Например, хотя задержка пересадки, возникающая при пересадке в северном направлении Зеленой линии на L'Enfant Plaza, коррелирует с задержками пересадки на той же станции для тех, кто пересаживается в южном направлении Желтой линии, а также с задержкой пересадки тех, кто пересаживается там в западном направлении Оранжевой линии, последние две не коррелируют друг с другом. Хотя на первый взгляд это может показаться нелогичным, это разумно, учитывая, что нет отношения origindestination, по которому пассажиры будут переключаться между этими службами, в отличие от вышеупомянутых взаимосвязанных задержек.

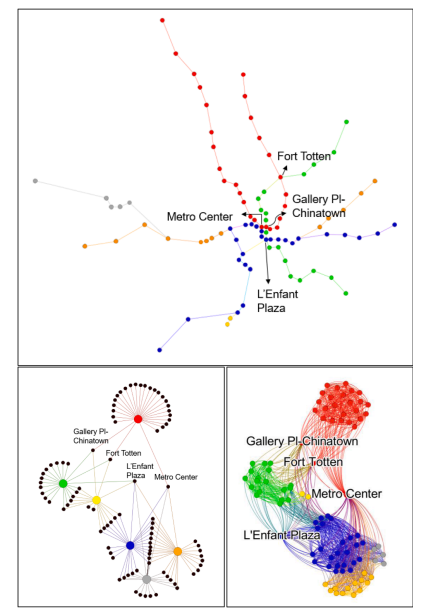


Рисунок 7 - Изображение сети Вашингтонского метрополитена в инфрапространстве (вверху), Переходное пространство (внизу слева) и служебное пространство (внизу справа).. Цвета узлов и линий соответствуют цвету обозначенной линии (см. рис. 1).

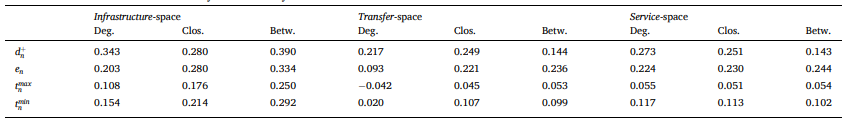


Таблица 1 - Корреляции между показателями информативности узлов и их центральности

3.4. Показатели информативности

Теперь мы переходим к извлечению показателей информативности, предложенных в разделе 2.2, для BN, полученных из наших данных о задержках пассажиров, как описано в предыдущем разделе. Мы вычисляем степень исходящего узла, ожидаемую прямую информативность, Нижнюю границу общей информативности и Верхнюю границу общей информативности для каждого узла в нашем графике BN. На рис. 6 снова показана конфигурация BN, где теперь цвет узла (задержка либо на начальной станции ожидания, либо на станции передачи) соответствует относительному значению соответствующего индикатора информативности. Для удобства использования положения всех узлов, которые здесь не соответствуют их географическому расположению, такие же, как на рис. 5, и они сохранены одинаковыми на всех графиках, включенных в рис. 6.

Индикатор степени исходящего узла напрямую отражает локальную связность узлов в BN. У него очень асимметричное распределение: несколько узлов, составляющих хабы в BN, имеют высокую степень, а большинство узлов имеют низкую степень, поскольку они подключены только к одному или двум другим узлам (рис. 6, вверху слева). Это означает, что задержки только на нескольких станциях являются прямой информацией для многих других станций. Это разумно, учитывая очень радиальную структуру сети тематических исследований и возможность инкапсулировать задержки без распространения их корреляции более чем на несколько станций.

Значения ожидаемой прямой информативности учитывают также метки ссылок, соединяющих с узлом в BN (уравнение (3)), а не только их количество (уравнение (2)). Заметно, что определенные узлы – как начальные, так и время ожидания передачи, – которые не подключены ко многим другим узлам, имеют высокое ожидаемое значение прямой информативности (рис. 6, вверху справа). Это обусловлено их высокой корреляцией с несколькими другими узлами, с которыми они коррелируют. Следовательно, это приводит к большему разнообразию значений индикаторов по всей сети. Найдено максимальное значение 0.53, в то время как теоретическое значение верхней границы равно единице, что означает, что один узел идеально коррелирует со всеми своими соседями.

Результаты для Нижней (уравнение (4)) и верхней (уравнение (5)) границ общей информативности в целом аналогичны (рис. 6, внизу слева и внизу справа соответственно). Высокие значения найдены для узлов, которые хорошо связаны по всему миру, поскольку они могут предоставлять прямо или косвенно (через другие узлы) информацию, которая имеет отношение ко многим другим узлам. Общее значение информативности для малого компонента в левом верхнем углу рисунков, которое соответствует начальным задержкам времени ожидания на средних станциях вдоль Красной линии против часовой стрелки, выделяется из-за клики, которую оно формирует. Относительно высокая сетчатость для этого компонента, который предлагает несколько путей в BN для получения информации для каждого узла, также, возможно, является причиной более выраженной разницы между нижними и верхними граничными значениями.

3.5. Взаимосвязь между показателями центральности и информативности

Мы представляем сеть тематических исследований в инфраструктурных, трансферных и сервисных пространствах (описанных в разделе 2.3), как показано на рис. 7. Инфраструктурное пространство близко соответствует знакомому представлению сети общественного транспорта, показанному на общедоступных картах. Пересадочное пространство состоит из двух наборов узлов: линий и станций. Последние подключены ко всем линиям, которые их обслуживают. Служебное пространство соединяет все пары остановок, до которых можно добраться без выполнения переноса, т. Е. Существует по крайней мере одна линия, обслуживающая обе остановки, и, следовательно, состоит из четко видимых кликов на основе строк.

Мы исследуем корреляции между тремя показателями центральности узла – степенью, промежуточностью и близостью - в каждом из трех графических представлений – инфраструктурных, трансферных и сервисных пространств – и предлагаемыми показателями информативности, извлеченными из BN, т. е. Степенью исходящего узла, ожидаемой прямой информативностью, нижней границей общей информативности и верхней границей общей информативности. Результаты представлены в таблице 1.

Можно заметить, что показатели топологии сети метро WMATA демонстрируют низкую или умеренную корреляцию с показателями информативности задержки пассажиров BN. Это указывает на то, что топологические свойства станции содержат лишь ограниченную информацию о том, в какой степени задержки, происходящие на этой станции, являются информативными относительно задержек, происходящих в течение того же временного интервала на других станциях в сети. Несмотря на это, некоторые умеренные корреляции потенциально могут быть полезны при отсутствии данных о полетах и пассажиропотоках. В частности, показатели, рассчитанные в рамках представления сети в инфраструктурном пространстве, и, в частности, центральность взаимосвязи, могут объяснить почти 40% наблюдаемой вариабельности. Наиболее всеобъемлющие показатели информативности, относящиеся к общей информативности, могут быть менее точно аппроксимированы строго топологическими показателями. Это подтверждает, что общесетевые соотношения задержек являются результатом сложных взаимодействий, которые невозможно легко аппроксимировать без адекватных источников данных.

4. Заключение

Степень одновременности задержек в транспортных сетях влияет на надежность и робастность сети. Мы количественно оцениваем объем информации, содержащейся в наблюдаемых задержках пассажиров в одном месте, о задержках пассажиров, наблюдаемых в других частях сети. Мы предлагаем серию показателей информативности, основанных на оценках байесовской сети. Кроме того, мы оцениваем, насколько хорошо предлагаемые показатели могут быть аппроксимированы показателями центральности соответствующих узлов.

Мы применяем предложенный метод к сети метро Вашингтона, округ Колумбия. Для анализа используется большой эмпирический набор данных, содержащий данные о движении пассажиров и поездов метро для практического применения. Полученный BN дает среднее значение RMSE 0,0092 мин. Как и следовало ожидать, задержки пассажиров на одной станции наиболее информативны для задержек, происходящих на соседних станциях того же направления линии. Кроме того, некоторые более удаленные друг от друга станции на той же линии или коридоре демонстрируют сильную корреляцию. Последнее, возможно, связано с зависимостью между диспетчеризацией, вызванной ограниченной инфраструктурой, а также с перемещениями пассажиров.

Задержки пассажиров на нескольких выбранных станциях напрямую информируют о задержках, происходящих на многих других станциях. Это говорит о том, что поставщики услуг могут использовать информацию с ограниченного набора станций для анализа состояния сети, что является возможным направлением будущих исследований. Результаты корреляционного анализа между показателями информативности и центральности позволяют предположить, что последние могут использоваться лишь в ограниченной степени в случае отсутствия данных о задержке пассажиров.

Корреляция не подразумевает причинно-следственной связи, и, следовательно, результаты не позволяют делать выводы о лежащих в основе наблюдаемых взаимосвязей между задержками пассажиров на станциях. Показатели информативности, предложенные в этом исследовании, и выявленные взаимосвязи должны быть дополнены аналитическими и имитационными моделями функционирования метрополитена, чтобы изучить взаимосвязи между (не) надежностью компонентов системы. Такие модели могут сыграть важную роль в изучении причин, приводящих к результату, с учетом общесетевых статистических свойств и взаимосвязей, описанных в этом исследовании. В будущих исследованиях такие модели могут быть откалиброваны таким образом, чтобы воспроизводить наблюдаемые случаи задержки и корреляции. Подход к транспортному моделированию будет включать комбинацию модели железнодорожного движения и модели распределения пассажиров, а также их калибровку и валидацию с использованием эмпирических данных, подобных тем, которые использовались в данном исследовании. Это проложит путь к потенциальным применениям, таким как устранение сбоев как в режиме реального времени, так и на тактическом уровне. Например, меры по снижению сбоев могут выиграть от включения информации о том, какие задержки могут быть инкапсулированы и которые, вероятно, возникнут в сочетании с задержками в других местах.

Представленные результаты действительны для сети тематических исследований и режима операций, использовавшихся в течение годичного периода анализа. Систематическое сравнение показателей информативности для ряда тематических исследований, которые различаются по своей топологии, использованию инфраструктуры, методам управления перегрузкой услуг и задержками, позволит сделать вывод о возможности передачи результатов, представленных в этом исследовании.

Дальнейшие исследования могут продвинуть предлагаемый метод вперед, включив в него также распространение временной задержки. BN может быть расширен, чтобы также включать задержки, зарегистрированные для разных временных срезов. Чтобы уменьшить вычислительные затраты, связанные с увеличением большого количества узлов, можно подключить каждый узел только к ограниченному числу непосредственно следующих друг за другом временных окон, в том числе между задержками на одной и той же станции в последовательных временных срезах для учета автокорреляции. В качестве альтернативы может быть разработана вероятностная сетевая модель, которая может фиксировать распространение сбоев с использованием механизмов, вдохновленных распространением эпидемий и слухов.