УДК 004.7
DOI: 10.26102/2310-6018/2025.50.3.026

Теоретические основы мониторинга изменений больших данных в крупномасштабных разреженных невзвешенных сетях с облачной обработкой

М.Д.Б. Аль-Имари[1], Д.В. Гетманская[2], О.Я. Кравец[2]✉, Д.В. Сотников[2]

[1]Казанский (Приволжский) федеральный университет, Казань, Российская Федерация
[2]Воронежский государственный технический университет, Воронеж, Российская Федерация

Резюме. Сети широко используются для представления интерактивных взаимосвязей между отдельными элементами в сложных системах больших данных, таких как облачный Интернет. Определяемые причины в системах могут приводить к резкому увеличению или уменьшению частоты взаимодействия в соответствующей сети, что позволяет выявлять такие определяемые причины, отслеживая уровень взаимодействия в сети. Один из методов обнаружения изменений заключается в том, что сначала между каждой парой узлов, которые взаимодействовали в течение заданного интервала времени, проводится ребро, чтобы создать сетевой граф. Затем топологические характеристики графа, такие как степень, близость и посредничество, могут рассматриваться как одномерные или многомерные данные для онлайн-мониторинга. Однако существующие методы статистического управления процессами (SPC) для невзвешенных сетей почти не учитывают ни разреженность сети, ни направление взаимодействия между двумя узлами сети, то есть парное взаимодействие. При исключении неактивных парных взаимодействий предложенная процедура оценки параметров обеспечивает более высокую согласованность при меньших вычислительных затратах, чем альтернативный вариант, когда сети являются крупномасштабными и разреженными. Разработанные на основе матричной вероятностной модели для описания направленных парных взаимодействий в рамках независимых от времени невзвешенных сетей больших данных с облачной обработкой матрицы значительно упрощают оценку параметров, эффективность которой повышается за счет автоматического исключения парных взаимодействий, которые на самом деле не происходят. Затем предложенная модель интегрируется в функцию многомерного распределения для онлайн-мониторинга уровня коммуникации в сети.

Ключевые слова: облачные вычисления, большие данные, изменения состояния сети, мониторинг в режиме реального времени, невзвешенные сети, парное взаимодействие, матричная вероятностная модель.

Для цитирования: Аль-Имари М.Д.Б., Гетманская Д.В., Кравец О.Я., Сотников Д.В. Теоретические основы мониторинга изменений больших данных в крупномасштабных разреженных невзвешенных сетях с облачной обработкой. *Моделирование, оптимизация и информационные технологии.* 2025;13(3). URL: https://moitvivt.ru/ru/journal/pdf?id=2004 DOI: 10.26102/2310-6018/2025.50.3.026

Theoretical foundations of monitoring big data changes in large-scale sparse unweighted networks with cloud processing

M.J.B. Al-Imari[1], D.V. Getmanskaia[2], O.Ja. Kravets[2]✉, D.V. Sotnikov[2]

[1]Kazan (Volga Region) Federal University, Kazan, the Russian Federation
[2]Voronezh State Technical University, Voronezh, the Russian Federation

© Аль-Имари М.Д.Б., Гетманская Д.В., Кравец О.Я., Сотников Д.В., 2025

Abstract. Networks are widely used to represent the interactive relationships between individual elements in complex big data systems, such as the cloud-based Internet. Determinable causes in these systems can lead to a significant increase or decrease in the frequency of interaction within the corresponding network, making it possible to identify such causes by monitoring the level of interaction within the network. One method for detecting changes is to first create a network graph by drawing an edge between each pair of nodes that have interacted within a specified time interval. The topological characteristics of the graph, such as degree, proximity, and mediation, can then be considered as one-dimensional or multidimensional data for online monitoring. However, the existing statistical process control (SPC) methods for unweighted networks almost do not take into account either the sparsity of the network or the direction of interaction between two network nodes, that is, pair interaction. By excluding inactive pair interactions, the proposed parameter estimation procedure provides higher consistency with lower computational costs than the alternative approach when the networks are large-scale and sparse. The matrices developed on the basis of a matrix probabilistic model for describing directed pair interactions within time-independent, unweighted big data networks with cloud processing significantly simplify parameter estimation, the effectiveness of which is increased by automatically eliminating pair interactions that do not actually occur. Then the proposed model is integrated into a multidimensional distribution function for online monitoring of the level of communication in the network.

Keywords: cloud computing, big data, network status changes, real-time monitoring, unweighted networks, pair interaction, matrix probability model.

For citation: Al-Imari M.J.B., Getmanskaia D.V., Kravets O.Ja., Sotnikov D.V. Theoretical foundations of monitoring big data changes in large-scale sparse unweighted networks with cloud processing. *Modeling, Optimization and Information Technology*. 2025;13(3). (In Russ.). URL: https://moitvivt.ru/ru/journal/pdf?id=2004 DOI: 10.26102/2310-6018/2025.50.3.026

Введение

Благодаря развитию информационных технологий все больше физических объектов могут быть связаны друг с другом с помощью датчиков или интеллектуальных устройств, образуя различные сетевые системы. Определенное количество взаимосвязанных «вещей» образует Интернет вещей (IoT), а взаимосвязь между группой людей формирует социальную сеть. В сети каждая вещь или человек могут взаимодействовать с другими для достижения общих целей. Например, центр обработки заказов Amazon оснащен несколькими тысячами роботов, которые взаимодействуют друг с другом при перемещении товаров. Сотрудники могут обмениваться электронными письмами со своими коллегами из того же отдела или компании для сотрудничества [1]. Примеры можно найти и во многих других областях, таких как нейронные сети [2] и машинные сети для интеллектуального производства [3]. Связи внутри этих сетей являются парными и направленными, поскольку они содержат передачу информации между парами отправителей и получателей.

Целью работы является создание теоретических основ мониторинга изменений больших данных в крупномасштабных разреженных невзвешенных сетях с облачной обработкой.

Материалы и методы

Состояние проблемы. Интерактивные сети обычно характеризуются изменчивостью, обусловленной их природой или внешними причинами. Первый тип изменчивости можно рассматривать как естественную эволюцию, в то время как второй, согласно [4], может привести к «изменениям». Такие изменения необходимы для обнаружения определенных аномалий в связанных системах. Например, незаконные действия, такие как сговор или мошенничество, могут спровоцировать резкое

увеличение частоты взаимодействия, например, в сети онлайн-аукционов [5], а неисправный компьютер с поддержкой Интернет сетей может перестать сообщать о своем состоянии другим машинам, что препятствует оптимальному планированию в интеллектуальном производстве [3]. В связи с этим в [6] отмечено, что выявление изменений в сети указывает на захватывающую новую область исследований, а в [7] предложено отслеживать уровень коммуникации в социальных сетях.

В последние годы появилось все больше методов статистического контроля процессов для выявления изменений в интерактивных сетях. Один из методов обнаружения изменений заключается в том, что сначала между каждой парой узлов, которые взаимодействовали в течение заданного интервала времени, проводится ребро, чтобы создать сетевой граф. Затем топологические характеристики графа, такие как степень, близость и посредничество, могут рассматриваться как одномерные или многомерные данные для онлайн-мониторинга. Среди работ можно выделить [8]. Другие исследователи использовали статистику сканирования подграфов или скользящее окно для разработки глобальных схем мониторинга [9]. Однако их предпочтительная статистика сканирования, основанная на максимальном количестве локальных показателей, может быть чувствительна только к очень значительным изменениям в сети [7]. Еще одна многомерная схема разработана в [10] для отслеживания соответствующих пропорций взаимных пар и транзитивных триад в сети. В целом, топологическое представление сетей часто может вводить в заблуждение, поскольку оно не было проверено с использованием реальных сетевых данных [11]. Результаты моделирования, полученные в [12], показывают, что некоторые типичные изменения в сети нельзя обнаружить с помощью методов мониторинга, основанных на средней степени связности и близости узлов.

Другой подход к схемам мониторинга основан на существующих вероятностных моделях, которые описывают механизм формирования ребер сети. Предположив, что каждое ребро является случайной величиной Бернулли, на которую влияют атрибуты узлов в дискретные моменты времени, в [13] предложено отслеживать результирующие невзвешенные сети с помощью тестов отношения правдоподобия, полученных из модели логистической регрессии. В [14] объединена обобщенная линейная модель и уравнение перехода между состояниями, чтобы предложить метод обнаружения изменений в невзвешенных сетях с потенциальной автокорреляцией ребер.

Однако вышеупомянутые подходы имеют два недостатка при моделировании и мониторинге невзвешенных сетей. Во-первых, они не учитывают направление парных взаимодействий, из-за чего связи внутри сети могут быть записаны с ошибками [7]. Например, когда один узел связывается с другим, текущие методы могут использовать только информацию о том, что между двумя узлами происходит контакт, независимо от его направления. Такая потеря информации может ограничить возможности методов мониторинга. Второй недостаток заключается в том, что методы статистического управления процессами (SPC) игнорируют разреженность невзвешенных сетей, которая вызвана тем, что некоторые узлы взаимодействуют лишь с частью остальных узлов в течение одного периода или даже постоянно. Разреженность в течение одного периода можно просто устранить, увеличив продолжительность этого периода или объединив наблюдения за несколькими последовательными периодами, в то время как устранение постоянной разреженности является сложной задачей, поскольку она возникает из-за пар узлов с нулевой вероятностью взаимодействия. Если не принять соответствующие меры, такая разреженность данных приведет к высоким вычислительным затратам при оценке методом максимального правдоподобия (MLE) и высокому риску неправильной спецификации модели.

В работе предполагаем, что невзвешенные сети не зависят от времени.

Моделирование управляемых невзвешенных сетей.

Постановка задачи. Рассмотрим сеть, состоящую из d узлов V_i, $i = 1 \dots d$. Обозначим $a_{ij} = 1$, если существует хотя бы один контакт от V_i к V_j в течение периода, и $a_{ij} = 0$ в противном случае. Тогда можно составить матрицу смежности A, в которой a_{ij} – элемент в i строке и j столбце. Условие $i \neq j$ вводится для исключения петель между узлами, и диагональные элементы в A должны быть соответственно равны нулю. В ориентированной сети $a_{ij} = a_{ji}$ не обязательно верно для A, которая в результате может быть асимметричной. Чтобы смоделировать ориентированные парные связи, определим $D_{ij} = (a_{ij}, a_{ji})$ как пару, а связь между узлами V_i и V_j как парное взаимодействие. Поскольку a_{ij} принимает значение 0 или 1, каждое парное взаимодействие может перейти в одно из четырех состояний: отключено с $D_{ij} = (0, 0)$, однонаправленно связано с $D_{ij} = (1, 0)$ или $D_{ij} = (0, 1)$ и двунаправленно связано с $D_{ij} = (1, 1)$ в течение одного периода. В этом смысле D_{ij} является случайной величиной в многопериодной эволюции сетей.

На распределение пар D_{ij} могут влиять внутренние характеристики узлов V_i и V_j, а также факторы окружающей среды. Например, сочетание ролей двух сотрудников определяет частоту их обмена электронными письмами, а суровые погодные условия могут снизить стремление всех физических устройств к взаимодействию друг с другом в мобильной сети. Для простоты изложения обозначим F как набор, содержащий все факторы, которые могут влиять на парные взаимодействия. Тогда получаем следующий результат:

$$Pr = \{D_{ij} = (a_{ij},\ a_{ji}),\ D_{kl} = (a_{kl},\ a_{lk}) | F\} = \\ = Pr\{D_{ij} = (a_{ij},\ a_{ji}) | F\} \times Pr\{D_{kl} = (a_{kl},\ a_{lk}) | F\},$$

где $i \neq k$ или $j \neq l$ и $Pr\{E|C\}$ указывает вероятность события E, зависящую от C.

Это означает, что любые две пары должны быть независимы друг от друга при заданном F. Эта условная независимость очевидна, поскольку любые две пары обычно определяются функцией F. На самом деле, такая условная независимость между парами была ранее использована при моделировании и мониторинге ненаправленных невзвешенных сетей.

Поскольку мы сосредоточены на мониторинге независимых от времени сетей, необходимо предположить, что факторы в F не будут меняться до тех пор, пока в связанных с ними системах не произойдет аномальное событие. Другими словами, парные вероятности $P_{ij}^{a_{ij}a_{ji}} = Pr\{D_{ij} = (a_{ij},\ a_{ji}) | F\}$ фиксируются, когда сетевая система работает нормально. Определяем $n_{ij}^{a_{ij}a_{ji}} = 1$ если $D_{ij} = (a_{ij},\ a_{ji})$ и $n_{ij}^{a_{ij}a_{ji}} = 0$ в противном случае. Тогда вектор наблюдений $n_{ij} = [n_{ij}^{00},\ n_{ij}^{10},\ n_{ij}^{01},\ n_{ij}^{11}]^T$ будет соответствовать многомерному распределению $MN(N,\ p_{ij})$ с $p_{ij} = [p_{ij}^{00},\ p_{ij}^{10}, p_{ij}^{01}, p_{ij}^{11}]^T$ и $1^T n_{ij} = N$, где N – количество равномерно распределенных временных интервалов. Для наглядности 25 выборок размером $N = 4$ последовательно формируются для представления сообщений между двумя пользователями, обозначенными как V_1 и V_2, в сети электронной почты крупного европейского исследовательского учреждения, описанной в [1]. Сначала мы объединяем все выборки в одну и делим ее на общий размер 100, чтобы получить оценку $p_{12} = [\hat{p}_{12}^{00}, \hat{p}_{12}^{10}, \hat{p}_{12}^{01}, \hat{p}_{12}^{11}]^T$.

Во-вторых, каждый образец можно использовать для вычисления статистики χ^2: $\chi^2(3) = \sum_{k=0}^{1} \sum_{l=0}^{1} \frac{(n_{12}^{kl} - N\hat{p}_{12}^{kl})^2}{N\hat{p}_{12}^{kl}}$, с 3 степенями свободы.

В-третьих, строим последовательные статистические графики, как показано на Рисунке 1, где верхний контрольный предел (UCL) и нижний контрольный предел (LCL) определяются в соответствии с ошибкой первого рода $\alpha = 0{,}05$. На Рисунке 1 показано, что значения статистики χ^2 в основном распределены случайным образом, что подтверждает утверждение о том, что наблюдения за парным взаимодействием подчиняются многомерному распределению.

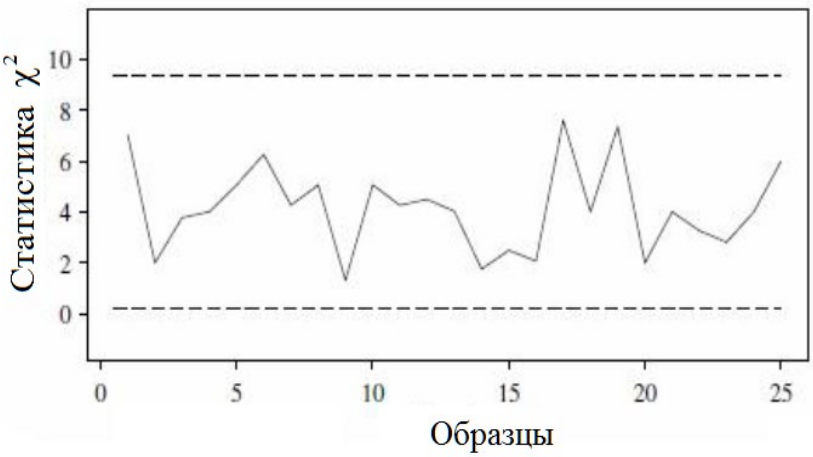

Рисунок 1 – Статистика χ^2 переписки двух человек по электронной почте с UCL = 9,35, LCL = 0,22

Figure 1 – Statistics of χ^2 correspondence between two people by e-mail with UCL = 9,35, LCL = 0,22

Результаты и обсуждение

Предлагаемая методология. Приведенный выше пример показывает, что в течение N временных интервалов каждое парное взаимодействие может быть представлено случайным вектором-столбцом, подчиняющимся многомерному распределению с размером выборки N. Благодаря независимости парных взаимодействий можно легко оснастить их отдельными схемами обнаружения изменений, основанными на многомерном распределении, и соответствующим образом объединить их для онлайн-мониторинга уровня коммуникации во всей сети. Однако этот метод игнорирует внутренние характеристики узлов и факторы окружающей среды F, что может привести к ограничению возможностей обнаружения в результате применения схемы мониторинга. В статье модель вероятности p^1, предложенная в [6], расширяется для создания более эффективного подхода.

В частности, склонность одного узла к контактам с другими можно назвать экспансивностью, а склонность других узлов к контактам с ним – его популярностью. Склонность всех узлов к общению друг с другом можно измерить с помощью взаимности. Таким образом, вероятность парного взаимодействия может определяться экспансивностью, популярностью и общей взаимностью этих двух узлов. Пусть $\alpha_i, \beta_i (i = 1, \ldots, d)$ соответственно обозначают распространенность и популярность узла i, p^1. Тогда модель вероятности для узлов V_i и V_j можно выразить как

$$ln\, p_{ij}^{00} = \lambda_{ij},$$
$$ln\, p_{ij}^{10} = \lambda_{ij} + \alpha_i + \beta_j + \gamma,$$
$$ln\, p_{ij}^{01} = \lambda_{ij} + \alpha_j + \beta_i + \gamma,$$
$$ln\, p_{ij}^{11} = \lambda_{ij} + \alpha_i + \alpha_j + \beta_i + \beta_j + 2\gamma + \rho,$$

где ρ – параметр взаимности, поскольку он присутствует только тогда, когда пара узлов находится во взаимном состоянии, λ_{ij} – необходимый параметр, как и во многих других регрессионных моделях, γ – общий объем отправленных и полученных вариантов. Ограничения $\sum_{i=1}^{d} \alpha_i = 0$, $\sum_{i=1}^{d} \beta_i = 0$ должны быть наложены для обеспечения идентифицируемости модели.

При статистической независимости $D_{ij} = (a_{ij},\ a_{ji})$, в [15] просто объединили все парные модели для характеристики всей сети. Поскольку наборы уравнений неудобны для использования, ниже они переписаны в матричной форме. Вспомним, что $p_{ij} = = \left[p_{ij}^{00},\ p_{ij}^{10}, p_{ij}^{01}, p_{ij}^{11}\right]^T$, и допустим, что $\theta_{ij} = \left[\alpha_i,\ \alpha_j,\ \beta_i,\ \beta_j, \gamma, \rho\right]^T$, тогда приведенные выше уравнения можно сформулировать в виде

$$ln\, p_{ij} = 1\lambda_{ij} + X\theta_{ij}, \text{ с } 1^T p_{ij} = 1,$$

с матрицей X, определенной следующим образом:

$$X = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

Поскольку λ_{ij} является функцией θ_{ij} в соответствии с $1^T p_{ij} = 1$, вероятность парного взаимодействия может определяться исключительно θ_{ij}. Становится сложно собрать все парные модели, поскольку компоненты θ_{ij} зависят от пары D_{ij}. Следовательно, для всех парных моделей необходимо общее параметрическое представление.

Предположим, что λ_{ij} временно исключена, а все остальные параметры модели выражены в виде вектора-столбца $\theta = [\alpha_1, \ldots,\ \alpha_{d-1}, \ldots,\ \alpha_d,\ \beta_1, \ldots,\ \beta_{d-1}, \beta_d, \gamma, \rho]^T$ размерности $(2d + 2) \times 1$, и можно составить матрицу Y_{ij} размерности $6 \times (2d + 2)$, удовлетворяющую условию $Y_{ij}\theta = \theta_{ij}$. Изначально для всех элементов в Y_{ij} может быть установлено значение 0. Чтобы получить α_i и β_i в качестве 1-го и 3-го элементов θ_{ij} соответственно, элементы в ячейках $(1, i)$ и $(3, d + i)$ таблицы Y_{ij} должны быть равны 1, а элементы в ячейках $(2, j)$ и $(4, d + j)$ также должны быть равны 1, чтобы α_j и β_j стали 2-м и 4-м элементами таблицы θ_{ij}. Кроме того, 5-й элемент γ и 6-й элемент ρ в θ_{ij} требуют, чтобы оба элемента в ячейке $(5, 2d + 1)$, $(6, 2d + 2)$ элемента Y_{ij} имели значение 1.

В соответствии с ограничениями идентифицируемости $\sum_{i=1}^{d} \alpha_i = 0$ и $\sum_{j=1}^{d} \beta_j = 0$, α_d, β_d можно вычислить, зная другие параметры. В связи с этим предположим, что $\theta = = [\alpha_1, \ldots,\ \alpha_{d-1},\ \beta_1, \ldots, \beta_{d-1}, \gamma, \rho]^T$, и спроектируем $(2d + 2) \times (2d)$ матрицу Z, удовлетворяющую $Z\theta = \theta$, в виде

$$Z = \begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & 0 & & & & & \\ -1 & \cdots & -1 & & & & & & \\ & & & 1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & 1 & & & \\ 0 & & & -1 & \cdots & -1 & & & \\ & & & & & & 1 & & \\ & & & & & & & 1 \end{bmatrix}.$$

Имеем $X\theta_{ij} = XY_{ij}Z\theta$. Пусть $XY_{ij}Z = Z_{ij}$, тогда модель можно преобразовать в

$$ln\, p_{ij} = 1\lambda_{ij} + Z_{ij}\theta. \quad (1)$$

Согласно ограничению $1^T p_{ij} = 1$, можно получить $\lambda_{ij} = -ln\left[1^T\, exp\left(Z_{ij}\theta\right)\right]$. Таким образом, все парные модели связаны с одним и тем же параметром θ, но различаются $4 \times (2d)$ матрицами Z_{ij}, которые можно спроектировать заранее, учитывая связанные с ними два узла. Для этого можно собрать все парные модели с учетом парной независимости для оценки параметров.

Условная оценка максимального правдоподобия. На практике сеть обычно состоит из большого количества узлов, что приводит к созданию огромных парных моделей. Оценка параметров модели с помощью отдельных итерационных шагов [4] потребует больших вычислительных мощностей. Благодаря матрицам, описанным выше, итоговую агрегированную модель можно выразить как функцию θ. Это позволяет легко вычислить MLE для θ в целом.

Напомним, что оценивается каждое парное взаимодействие в конце заданного временного интервала. Пара узлов V_i и V_j должна быть одной из $D_{ij} = (0,0)$, $D_{ij} = (1,0)$, $D_{ij} = (0,1)$, $D_{ij} = (1,1)$. Как описано в разделе «Постановка задачи», наблюдаемые пары можно преобразовать в случайные векторы $n_{ij} = \left[n_{ij}^{00},\, n_{ij}^{10},\, n_{ij}^{01},\, n_{ij}^{11}\right]^T$, подчиняющиеся многомерному распределению $MN\left(N, p_{ij}\right)$ с размером выборки $N = 1^T n_{ij}$ и $p_{ij} = \left[p_{ij}^{00}, p_{ij}^{10},\, p_{ij}^{01}, p_{ij}^{11}\right]^T$. Так, если наблюдается $D_{ij} = (1,0)$ в течение одного временного интервала, вектор подсчета может быть обозначен как $n_{ij} = [0,1,0,0]^T$. Функция распределения определяет логарифмическую вероятность n_{ij} как:

$$l_{ij}(\theta) = n_{ij}^{00}\, ln\, p_{ij}^{00} + n_{ij}^{10}\, ln\, p_{ij}^{10} + n_{ij}^{01}\, ln\, p_{ij}^{01} + n_{ij}^{11}\, ln\, p_{ij}^{11} = n_{ij}^T\, ln\, p_{ij} =$$
$$= n_{ij}^T\left(-1\, ln\left[1^T\, exp\left(Z_{ij}\theta\right)\right] + Z_{ij}\theta\right) = -N\, ln\left[1^T\, exp\left(Z_{ij}\theta\right)\right] + n_{ij}^T Z_{ij}\theta.$$

Благодаря независимости парных моделей можно сложить $l_{ij}(\theta)$, чтобы получить полную функцию логарифмической вероятности для сети. Т.к. $p_{ij}^{10} = p_{ji}^{01}$ и $n_{ij}^{10} = n_{ji}^{01}$, функцию логарифмической вероятности можно выразить как $l(\theta) = \sum_{i<j} l_{ij}(\theta)$.

Согласно вышеупомянутым настройкам, эта функция содержит $2d$-параметры для сети с d узлами. Легко получить производные этой функции правдоподобия по θ.

Однако многие узлы в крупномасштабной сети могут вообще не взаимодействовать друг с другом, что приводит к разреженности сети. Предположим, что между V_i и V_j не происходит никаких взаимодействий, когда и $i < j$ и $i, j = 1, 2, \ldots, d$, есть $p_{ij}^{00} = Pr\{D_{ij} = (0,0)\} = 1$ и наблюдается $n_{ij} = [N,0,0,0]^T$. Поскольку $\theta = [\alpha_1, \ldots, \alpha_{d-1}, \beta_1, \ldots, \beta_{d-1}, \gamma, \rho]^T$, соответствующую $2d \times 2d$ информационную

матрицу можно обозначить как $\hat{I}(\theta) = \frac{-N^{-1}\partial l^2(\theta)^T}{\partial\theta\partial\theta}$, где $\hat{I}_{kl}(\theta)$ – элемент в строке k и столбце l. С помощью алгебраических преобразований можно получить:

Утверждение 1. Если для узлов V_i и V_j с $i < j$ и $i, j = 1, 2, \ldots, d$ выполнено $p_{ij}^{00} = = 1$, то

$$\hat{I}_{ij}(\theta) = -\frac{1}{N}\sum_{k=1}^{d-1} I\{n_{kd}^{00}<\mathrm{N}\}\frac{\partial^2 l_{kd}(\theta)}{\partial\alpha_d^2},$$

$$\hat{I}_{i(j+d-1)}(\theta) = \hat{I}_{(i+d-1)j}(\theta) = -\frac{1}{N}\sum_{k=1}^{d-1} I\{n_{kd}^{00}<\mathrm{N}\}\frac{\partial^2 l_{kd}(\theta)}{\partial\alpha_d\partial\beta_d},$$

$$\hat{I}_{(i+d-1)(j+d-1)}(\theta) = -\frac{1}{N}\sum_{k=1}^{d-1} I\{n_{kd}^{00}<\mathrm{N}\}\frac{\partial^2 l_{kd}(\theta)}{\partial\beta 2_d},$$

где $\alpha_d = -(\alpha_1 + \cdots + \alpha_{d-1})$, $\beta_d = -(\beta_1 + \cdots + \beta_{d-1})$ и $I\{E\} = 1$, если выполняется условие E и $I\{E\} = 0$ в противном случае. Соответствующие симметричные элементы в $\hat{I}(\theta)$ могут быть вычислены аналогичным образом.

Доказательство. Каждая запись в $\hat{I}(\theta)$, связанная с v_i и v_j или математически с α_i, α_j, β_i, β_j, рассчитывается на основе $l_{ij}(\theta)$ и $l_{kd}(\theta)$, $k = 1, \ldots, d-1$. Когда $Pr\{D_{ij} = (0,0)\} = 1$, производная второго порядка от $l_{ij}(\theta)$ по четырем параметрам будет равна 0. Поскольку $\frac{\partial l_{kd}(\theta)}{\partial\alpha_i} = \frac{\partial l_{kd}(\theta)}{\partial\alpha_j} = \frac{-\partial l_{kd}(\theta)}{\partial\alpha_d}$ и $\frac{\partial l_{kd}(\theta)}{\partial\beta_i} = \frac{\partial l_{kd}(\theta)}{\partial\beta_j} = \frac{-\partial l_{kd}(\theta)}{\partial\beta_d}$, результаты из Утверждения 1 можно получить напрямую. При этом n_{kd}^{00}<N индуцируется $Pr\{D_{kd} = (0,0)\} \neq 1$.

Это говорит о том, что при $p_{ij}^{00} = 1$ записи, связанные с V_i и V_j в $\hat{I}(\theta)$, будут зависеть не от пар D_{ij}, а от пар, связанных с V_d. Другими словами, наблюдаемые значения $n_{ij} = [N, 0, 0, 0]^T$ дают мало информации для оценки θ, и эта малозначимая информация обычно используется всеми α_i и всеми β_j соответственно. Из функции логарифмической вероятности $l_{ij}(\theta)$ также следует, что параметр θ_{ij} для узлов V_i и V_jедва ли можно оценить на основе $n_{ij} = [N, 0, 0, 0]^T$. Как только задействуются неактивные пары, параметры, рассчитанные на основе активных парных взаимодействий, будут соответствовать наблюдению $n_{ij} = [N, 0, 0, 0]^T$. Это приведет к тому, что оценка будет отклоняться от истинного значения и, следовательно, снизится точность оценки параметров. Чтобы сэкономить вычислительное время и избежать неправильной спецификации модели, особенно для крупномасштабных разреженных сетей, наблюдаемые значения $n_{ij} = [N, 0, 0, 0]^T$ не будут учитываться в соответствии с [6]. Соответствующая условная функция логарифмической вероятности запишется как

$$l_c(\theta) = \sum_{n_{ij}^{00}<\mathrm{N}}^{i<\mathrm{j}} l_{ij}(\theta),$$

где нижний индекс «c» означает «условная». Для большей наглядности количество активных парных взаимодействий, то есть взаимодействий с n_{ij}^{00}<N для узла $V_i(i = 1, \ldots, d)$, можно обозначить как $n_i^* = \sum_{k=1}^{d} I\{n_{ik}^{00}<\mathrm{N}\}$. Очевидно, что при фиксированном d чем больше n_i^*, тем эффективнее будет MLE для θ. Чтобы обеспечить

уникальность решения, можно объединить все Z_{ij}, связанные с активными парными узлами, в вертикальную матрицу большего размера $B = \{Z_{ij}\}_{i<j}$ с размерностью $\left(4\sum_{i=1}^{d} n_i^*\right) \times (2d)$ и предположить, что B имеет ранг $2d$. Тогда для получения θ можно использовать алгоритм Ньютона-Рафсона:

$$\theta^{(\xi+1)} = \theta^{(\xi)} + \frac{1}{N}\left[\sum_{n_{ik}^{00}<N}^{i<j} Z_{ij}^T \Lambda_{ij}^{(\xi)} Z_{ij}\right]^{-1} \sum_{n_{ik}^{00}<N}^{i<j} Z_{ij}^T \left(n_{ij} - N p_{ij}^{(\xi)}\right),$$

где ξ – итерация, N – размер выборки и $\Lambda_{ij}^{(\xi)} = diag\left(p_{ij}^{(\xi)}\right) - p_{ij}^{(\xi)} p_{ij}^{(\xi)T}$ с $p_{ij}^{(\xi)} = = exp\left(Z_{ij}\theta^{(\xi)} - 1\, ln\left(1^T\, exp\left(Z_{ij}\theta^{(\xi)}\right)\right)\right)$. Здесь $diag\left(p_{ij}^{(\xi)}\right)$ – диагональная квадратная матрица с диагональными элементами из $p_{ij}^{(\xi)}$.

Результаты моделирования показывают [16], что алгоритм требует всего несколько итераций для сходимости, а также обеспечивает высокую точность оценки. Следует отметить, что предложенный метод может работать не очень хорошо, когда масштаб параметров становится слишком большим из-за особенностей метода максимального правдоподобия. В таких случаях необходимо использовать процедуры выбора переменных, например, оценку максимального правдоподобия с штрафами.

Заключение

Статья посвящена сетям, не зависящим от времени, и может быть расширена до временных динамических сетей с последовательной зависимостью или сезонностью. Предлагается матричная вероятностная модель для описания направленных парных взаимодействий в рамках независимых от времени невзвешенных сетей больших данных с облачной обработкой. Разработанные матрицы значительно упрощают оценку параметров, эффективность которой повышается за счет автоматического исключения парных взаимодействий, которые на самом деле не происходят. Затем предложенная модель интегрируется в функцию многомерного распределения, чтобы получить две схемы, основанные на методе GLRT, для онлайн-мониторинга уровня коммуникации в сети. Одно из расширений работы может быть реализовано с помощью процедур выбора переменных, которые позволяют решить проблему мониторинга сетей с более высокой или сложной разреженностью.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Paranjape A., Benson A.R., Leskovec J. Motifs in Temporal Networks. In: *WSDM '17: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, 06–10 February 2017, Cambridge, United Kingdom.* New York: Association for Computing Machinery; 2017. P. 601–610. https://doi.org/10.1145/3018661.3018731
2. Li B., Solea E. A Nonparametric Graphical Model for Functional Data with Application to Brain Networks Based on fMRI. *Journal of the American Statistical Association.* 2018;113(524):1637–1655. https://doi.org/10.1080/01621459.2017.1356726
3. Yang H., Kumara S., Bukkapatnam S.T.S., Tsung F. The Internet of Things for Smart Manufacturing: A Review. *IISE Transactions.* 2019;51(11):1190–1216. https://doi.org/10.1080/24725854.2018.1555383

4. Zou N., Li J. Modeling and Change Detection of Dynamic Network Data by a Network State Space Model. *IISE Transactions.* 2017;49(1):45–57. https://doi.org/10.1080/0740817X.2016.1198065
5. Pandit Sh., Chau D.H., Wang S., Faloutsos Ch. Netprobe: A Fast and Scalable System for Fraud Detection in Online Auction Networks. In: *WWW '07: Proceedings of the 16th International Conference on World Wide Web, 08–12 May 2007, Banff, Alberta, Canada*. New York: Association for Computing Machinery; 2007. P. 201–210. https://doi.org/10.1145/1242572.1242600
6. McCulloh I., Carley K.M. Detecting Change in Longitudinal Social Networks. *Journal of Social Structure.* 2011;12(1). https://doi.org/10.21307/joss-2019-031
7. Woodall W.H., Zhao M.J., Paynabar K., Sparks R., Wilson J.D. An Overview and Perspective on Social Network Monitoring. *IISE Transactions.* 2017;49(3):354–365. https://doi.org/10.1080/0740817X.2016.1213468
8. Hosseini S.S., Noorossana R. Performance Evaluation of EWMA and CUSUM Control Charts to Detect Anomalies in Social Networks Using Average and Standard Deviation of Degree Measures. *Quality and Reliability Engineering International.* 2018;34(4):477–500. https://doi.org/10.1002/qre.2267
9. Marchette D. Scan Statistics on Graphs. *WIREs Computational Statistics.* 2012;4(5):466–473. https://doi.org/10.1002/wics.1217
10. Perry M.B. An EWMA Control Chart for Categorical Processes with Applications to Social Network Monitoring. *Journal of Quality Technology.* 2019;52(2):182–197. https://doi.org/10.1080/00224065.2019.1571343
11. Goldenberg A., Zheng A.X., Fienberg S.E., Airoldi E.M. A Survey of Statistical Network Models. *Foundations and Trends® in Machine Learning*. 2010;2(2):129–233. https://doi.org/10.1561/2200000005
12. Dong H., Chen N., Wang K. Modeling and Change Detection for Count-Weighted Multilayer Networks. *Technometrics.* 2020;62(2):184–195. https://doi.org/10.1080/00401706.2019.1625812
13. Azarnoush B., Paynabar K., Bekki J., Runger G. Monitoring Temporal Homogeneity in Attributed Network Streams. *Journal of Quality Technology.* 2016;48(1):28–43. https://doi.org/10.1080/00224065.2016.11918149
14. Gahrooei M.R., Paynabar K. Change Detection in a Dynamic Stream of Attributed Networks. *Journal of Quality Technology.* 2018;50(4):418–430. https://doi.org/10.1080/00224065.2018.1507558
15. Holland P.W., Leinhardt S. An Exponential Family of Probability Distributions for Directed Graphs. *Journal of the American Statistical Association.* 1981;76(373):33–50. https://doi.org/10.1080/01621459.1981.10477598
16. Гетманская Д.В. Исследование парных взаимодействий в мониторинге невзвешенных ненаправленных сетей. *Системы управления и информационные технологии.* 2025;(2-1):30–36.
Hetmanskaya D.V. Investigation of Pair Interactions in Monitoring Unweighted Undirected Networks. *Sistemy upravleniya i informatsionnye tekhnologii.* 2025;(2-1):30–36. (In Russ.).

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Аль-Имари Мустафа Джафар Бакер, аспирант, Казанский (Приволжский) федеральный университет, Казань, Российская Федерация.
e-mail: csit@bk.ru

Al-Imari Mustafa Jaafar Baqer, postgraduate, Kazan (Volga Region) Federal University, Kazan, the Russian Federation.

Гетманская Диана Викторовна, аспирантка, Воронежский государственный технический университет, Воронеж, Российская Федерация.
e-mail: csit@bk.ru

Diana V. Getmanskaia, postgraduate, Voronezh State Technical University, Voronezh, the Russian Federation.

Кравец Олег Яковлевич, доктор технических наук, профессор, профессор кафедры автоматизированных и вычислительных систем, Воронежский государственный технический университет, Воронеж, Российская Федерация.
e-mail: csit@bk.ru
ORCID: 0000-0003-0420-6877

Oleg Ja. Kravets, Doctor of Engineering Sciences, Full Professor, Professor at the Department of Automated and Computational Systems, Voronezh State Technical University, Voronezh, the Russian Federation.

Сотников Дмитрий Владимирович, аспирант, Воронежский государственный технический университет, Воронеж, Российская Федерация.
e-mail: csit@bk.ru

Dmitry V. Sotnikov, postgraduate, Voronezh State Technical University, Voronezh, the Russian Federation.

Статья поступила в редакцию 25.06.2025; одобрена после рецензирования 14.07.2025; принята к публикации 28.07.2025.

The article was submitted 25.06.2025; approved after reviewing 14.07.2025; accepted for publication 28.07.2025.