

КИМАН А. М. к.т.н., доцент

ПРОХОРЧЕНКО Г. О., к.т.н., доцент

КРАВЧЕНКО М. А., к.т.н.

(Український державний університет залізничного транспорту)



## Розроблення методу інтеграції операційних даних залізничної мережі з графовою структурою плану формування поїздів

**Анотація.** У статті розроблено метод інтеграції операційних даних функціонування залізничної мережі з графовою структурою плану формування поїздів. Метод спрямований на формування інформаційної основи для аналізу та прогнозування затримок поїздопотоків з урахуванням просторової структури транспортної системи. Для дослідження використано операційні дані АТ «Укрзалізниця» про навантаження вагонів різних номенклатур вантажів та інформацію про простоті кинутих поїздів на станціях полігона мережі регіональної філії «Одеська залізниця». Розроблено процедуру агрегування даних за групами вантажів і видами вагонопотоків і перетворено їх у показники поїздо-годин затримок. Запропоновано алгоритм прив'язки станцій навантаження і станцій затримки до опорних станцій плану формування поїздів із використанням графової моделі залізничної мережі. У результаті сформовано інтегрований набір даних, що поєднує параметри навантаження, маршрути руху вагонопотоків і показники затримок поїздів. Отриманий набір даних може бути використаний як навчальна вибірка для побудови математичних моделей аналізу і прогнозування затримок поїздопотоків у залізничній мережі.

**Ключові слова:** залізнична мережа, план формування поїздів, графова модель, вагонопотоки, затримки поїздів, прогнозування.

### Вступ.

Одним із найбільш чутливих індикаторів ефективності функціонування вантажної залізничної системи є затримки поїздів, що відображають перевантаження окремих елементів мережі та дисбаланс у розподілі вагонопотоків [1]. Традиційні методи аналізу таких процесів здебільшого базовані на локальних показниках роботи залізничних станцій або окремих дільниць, через що не можна повною мірою врахувати просторову структуру руху вагонопотоків у масштабі всієї мережі. У зв'язку з цим особливої актуальності набуває використання мережевих підходів, які дають змогу розглядати залізничну систему як складну взаємопов'язану структуру та досліджувати взаємозалежності між її елементами.

Основним нормативним документом, що регулює систему організації вагонопотоків у поїзди, є план формування поїздів (ПФП) [2], який визначає схему консолідації вагонів у поїзди, маршрути їх просування та розподіл технологічних функцій між станціями. Із позицій теорії складних мереж ПФП може бути інтерпретований як графова модель залізничної системи, що можна використати для математичного моделювання прогнозованих затримок поїздопотоків на мережі. Однак для практичного застосування мережевих моделей потрібно розв'язати проблему узгодження різнорідних операційних даних із структурою системи організації вагонопотоків. Зокрема, дані про навантаження вагонів, маршрути їх переміщення та фактичні затримки поїздів

сформовані в різних інформаційних масивах і не мають прямої відповідності з графовою структурою ПФП. Це ускладнює використання таких даних для побудови математичних моделей прогнозування.

У зв'язку з цим актуальним науковим завданням є розроблення методу інтеграції операційних даних залізничної мережі з графовою структурою плану формування поїздів, що забезпечує формування узгоджених масивів даних для подальшого моделювання та прогнозування затримок поїздів. Реалізація такого підходу допомагає врахувати просторову структуру руху вагонопотоків, зменшити розмірність вихідних даних і сформувати навчальні вибірки для побудови математичних моделей прогнозування затримок поїздопотоків на мережі.

### Аналіз останніх досліджень і публікацій.

Проблема дослідження затримок поїздів є одним із складних завдань. Розв'язуючи цю проблему, складно визначити релевантний набір даних, які можуть бути використані для побудови моделей, що описують складні процеси утворення заторів. Сучасні інформаційні системи залізничного транспорту генерують значні масиви різнорідних операційних даних, що відображають параметри навантаження, маршрути руху вагонопотоків, технологічні операції на станціях і фактичні затримки поїздів. Однак ефективне використання цих даних потребує їх узгодження, агрегування та виявлення статистичних залежностей між характеристиками перевізного процесу і виникненням затримок.

© КИМАН А. М., ПРОХОРЧЕНКО Г. О., КРАВЧЕНКО М. А., 2026

## ІНФОРМАЦІЙНО-КЕРУЮЧІ СИСТЕМИ НА ЗАЛІЗНИЧНОМУ ТРАНСПОРТІ

У роботах, присвячених аналізу та прогнозуванню затримок поїздів, можна виділити кілька основних напрямів використання даних і методів їх агрегування та обробки. У більшості досліджень прогнозування затримок ґрунтовано на використанні історичних даних про рух поїздів, які містять інформацію про плановий і фактичний час прибуття і відправлення поїздів [3], величину відхилення від графіка руху [4], а також попередні затримки на станціях або дільницях [5, 6]. Такі дані формують часові ряди експлуатаційних показників, на основі яких будують статистичні, стохастичні або моделі прогнозування на основі машинного навчання. Основною особливістю цього підходу є подання транспортного процесу як послідовності подій у часі, що описують проходження поїздами станцій мережі. Водночас у значній частині досліджень просторові взаємозв'язки між елементами залізничної мережі враховані обмежено [6, 7] або лише опосередковано [8].

У дослідженнях українських учених значну увагу приділено удосконаленню математичних моделей розрахунку плану формування поїздів. Зокрема, у роботі [9] запропоновано уточнення моделі ПФП через урахування впливу завантаженості сортувальних станцій на витрати вагоно-годин. Для пошуку оптимальних варіантів плану формування поїздів використано дані, що мають технологічний і плановий характер, зокрема показники організації вагонопотоків, технологічні параметри роботи станцій, експлуатаційні витрати. Однак у роботі основну увагу зосереджено на оптимізації процесу формування поїздів і витрат вагоно-годин, тоді як задачу інтеграції операційних даних із системою організації вагонопотоків на рівні залізничної мережі безпосередньо не розглянуто. У дослідженні [10] для прогнозування обсягів пасажирських перевезень застосовано агреговані статистичні дані функціонування залізничної транспортної системи. До таких даних належать показники обсягів перевезених пасажирів, фактори, що впливають на формування попиту на перевезення, а також узагальнені характеристики транспортної системи і маршрутів перевезень. На основі цих показників сформовано вхідний набір параметрів прогнозної моделі пасажиропотоку. Використання саме агрегованих статистичних даних допомагає встановити залежності між факторами попиту і обсягами перевезень і застосувати їх для прогнозування пасажирських потоків у системі залізничного транспорту. Однак просторові характеристики пасажиропотоків не враховані. У роботі [1] для аналізу поширення затримок поїздів у залізничній мережі використано модифіковану епідеміологічну SIR-модель, у якій емпіричні дані про затримки агреговані за дільницями мережі з урахуванням пріоритетів поїздів і мережевого ефекту їхньої взаємодії. При цьому питання інтеграції різнорідних операційних даних перевізного процесу, зокрема даних про навантаження вагонів, простій

поїздів і їхню взаємодію з технологічною структурою плану формування поїздів, безпосередньо не розглянуто. Це зумовлює необхідність розроблення методів формування та інтеграції операційних даних залізничної мережі з її графовою структурою, що дає змогу забезпечити більш повний просторово-часовий опис процесів виникнення та поширення затримок поїздів. У дослідженні [11] проаналізовано вплив обсягів перевезень і щільності транспортної роботи на економічну ефективність функціонування залізничної системи України; виконано макрорівневий аналіз даних із виявленням взаємозалежностей і не розглянуто безпосередньо експлуатаційні процеси руху поїздів. У роботі [12] виконано порівняльний аналіз різних схем організації перевезень, що дає змогу оцінити їхній вплив на ефективність використання вагонного парку та час доставлення вантажів.

Останні дослідження дедалі більше уваги приділяють врахуванню просторової структури залізничної системи [13, 14]. У роботі [13] на основі мережево-орієнтованого підходу, де залізниця в кожній часовій мітці розглянута як окремий граф, різні об'єкти, такі як поїзди і станції, є вузлами. У роботі [14] запропонований метод агрегації даних про обсяги перевезень температурно-чутливих вантажів у залізничній логістичній системі за просторовими об'єктами транспортної мережі (містами) і часовими інтервалами спостереження. Отримані агреговані показники використовують як вхідні параметри графової нейронної мережі. Усе більше досліджень спираються на методи інтеграції даних із просторовими структурами організації перевезень.

Отже, аналіз наукових досліджень показує, що важливим етапом побудови, що описують складні процеси утворення заторів, є формування узгоджених масивів вхідних даних, які мають відображати як часові характеристики експлуатаційних процесів, так і просторову структуру транспортної мережі. Разом із тим у більшості існуючих робіт недостатньо уваги приділено інтеграції операційних даних про формування вагонопотоків, навантаження та маршрути їх просування з просторовою структурою організації вагонопотоків у поїзди. Зокрема, можливість використання графової структури плану формування поїздів для формування просторово узгоджених наборів даних для вивчення заторів на залізниці залишаються недостатньо дослідженими.

#### **Визначення мети та завдання дослідження.**

Метою роботи є розроблення методу інтеграції операційних даних функціонування залізничної мережі з графовою структурою плану формування поїздів для формування інформаційної основи аналізу та прогнозування затримок поїздопотоків у залізничній системі.

Для досягнення поставленої мети потрібно вирішити такі завдання:

– розробити підхід щодо агрегування та перетворення операційних даних залізничної мережі,

## ІНФОРМАЦІЙНО-КЕРУЮЧІ СИСТЕМИ НА ЗАЛІЗНИЧНОМУ ТРАНСПОРТІ

зокрема інформації про обсяги навантаження та випадки простою кинутих поїздів, у показники поїздо-годин затримок, що характеризують порушення технології перевізного процесу;

– сформувати метод інтеграції отриманих експлуатаційних показників із графовою моделлю плану формування поїздів для подальшого аналізу та прогнозування затримок поїздопотоків з урахуванням просторової структури залізничної мережі.

**Викладення основного матеріалу.** У статті запропоновано метод інтеграції операційних даних залізничної мережі з графовою структурою плану формування поїздів. Цей метод агрегує дані про навантаження, маршрут руху вагонів із розподілом за трьома родами вантажу і видом вагонопотоку, і значення поїздо-годин затриманих поїздів за станціями ПФП для прогнозування затримок.

Для проведення цього дослідження використано операційні дані АТ «Укрзалізниця» про навантаження зернових вантажів, будматеріалів та інших вантажів, крім вантажів гірничо-металургійного комплексу, за період з 1 вересня 2019 року по січень 2020 року, які містять дані з 110567 записів із датою і часом про обсяги навантаження вагонів за станціями відправлення і призначення, кодами вантажів (рис. 1), а також дані з 7682 записів із датою і часом про кількість і тривалість простою кинутих поїздів на станціях полігона мережі регіональної філії «Одеська залізниця» (довідка форми «Детальна довідка про затримку поїздів по станціям дирекції Одеська залізниця більше 12 годин за період звітних дат») (рис. 2).

dataload.xlsx

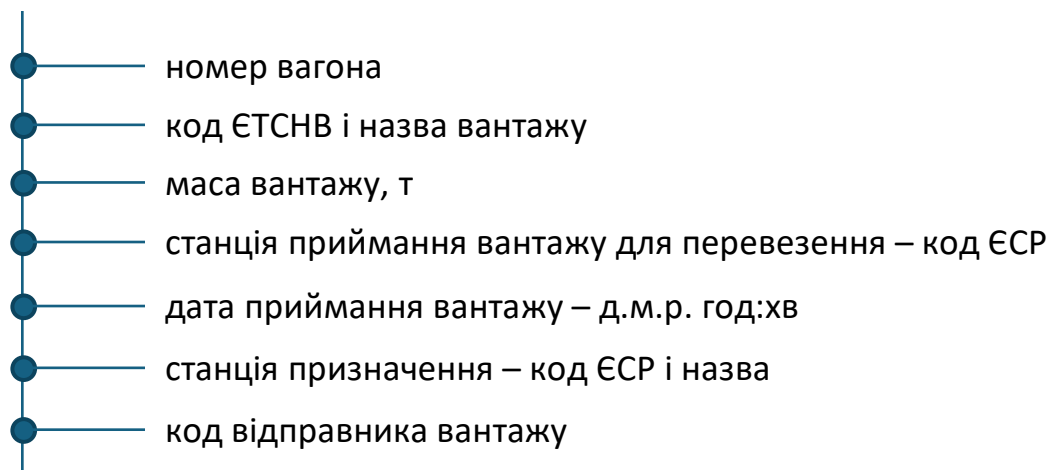


Рис. 1. Структура масиву даних про завантаження вагонів на залізничному полігоні

datadelay.xlsx

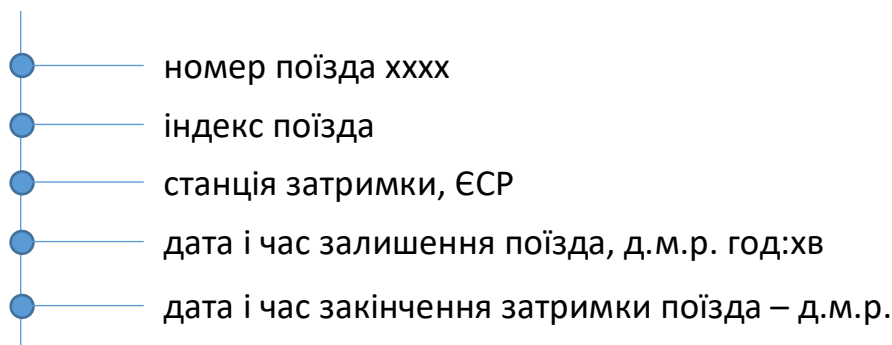


Рис. 2. Структура масиву даних про кинуті поїзди на залізничному полігоні

На першому етапі метод передбачає агрегацію вантажів за кодом ЄТСНВ – Єдина Тарифно-Статистична Номенклатура Вантажів [15]. Кодують вантаж згідно з ЄТСНВ за шестизначним кодом, де перші дві цифри коду позначають порядковий номер

групи, що визначає належність до певної номенклатури вантажів. На цій основі було виокремлено три узагальнені групи вантажів: агропродукція – товари, що належать до класів з кодами 01 та 02 (сільськогосподарська продукція), 50-60 (продукти харчування, напої, товари широкого

## ІНФОРМАЦІЙНО-КЕРУЮЧІ СИСТЕМИ НА ЗАЛІЗНИЧНОМУ ТРАНСПОРТІ

вжитку); вантажі, що належать до будівельних матеріалів (будмат) – вантажі з кодами 23 (матеріали мінерально-будівельні природні, зола, шлаки, крім гранульованих металургійних для переплавки), 25 (матеріали стінові покрівельні), 26 (матеріали, вироби будівельні промислового виробництва, крім стінових і покрівельних); інші вантажі – усі решта позицій класифікатора, які не належать до вищезазначених груп. Таке групування ґрунтовано на розподілі загального обсягу перевезеного вантажу за номенклатурою, де домінують будматеріали, а також

зернові та олійні культури. Розроблений алгоритм виконує функцію групування вантажів, яка, аналізуючи перші дві цифри коду ЄТСНВ, відносить запис до узагальнених груп «агро», «будматеріали» або «інші». Виконують агрегувальний SQL-запит із групуванням за номенклатурою вантажу, групою вантажу, станціями відправлення та прибуття і датою, який обчислює сумарну вагу у тоннах і кількість вагонів. Результати переносять у вихідний файл Excel за назвою results.xlsx (рис. 3).

result.xlsx

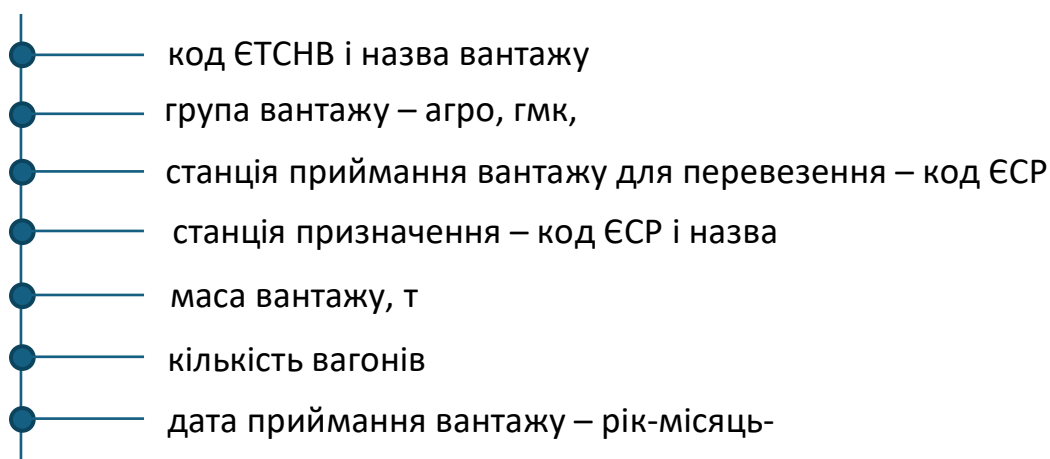


Рис. 3. Структура масиву даних після групування вантажів

Окрім типу вантажу, для опису різної складності технологічного процесу перевезення запропоновано врахувати вид відправлення [16], який визначають за кількістю вагонів, навантажених одночасно на одній станції. Агрегування даних за типом відправлення виконано через поділ навантаження на маршрутні відправки, що ідентифікують за умови одночасного приймання для перевезення з однаковою станцією призначення та отримувачем у понад сорок вагонів, та вагонні відправки, до яких належать групи вагонів із кількістю вагонів до сорока включно. Поділ відправлень на маршрутні та вагонні дає змогу поділити вагонопотоки в ПФП на транзитні з переробкою (тр з/пер) і транзитні без переробки (тр б/пер), що відображає різний рівень складності технологічних операцій на станціях мережі та потенційну ймовірність виникнення затримок.

Запропонований підхід дає змогу зберегти інтерпретацію даних про кількісні та технологічні особливості впливу на перевантаження залізничної мережі і водночас зменшити їхню вимірність для побудови прогнозу моделі.

Для врахування графової структури Плану формування поїздів (ПФП) запропоновано інтегрувати дані про навантаження вагонів і дані про кинуті поїзди з графом ПФП, що описує систему організації вагонопотоків у поїзди на 2018-2019 рік. Оцифрований граф мережі призначень ПФП має для кожної вершини назву станції та унікальний номер станції в мережі (код ЄСР). У межах цього дослідження використано графову структуру ПФП за найменуванням G2019. На рис. 4 наведена концептуальна схема інтеграції даних про навантаження вагонів із графом ПФП.

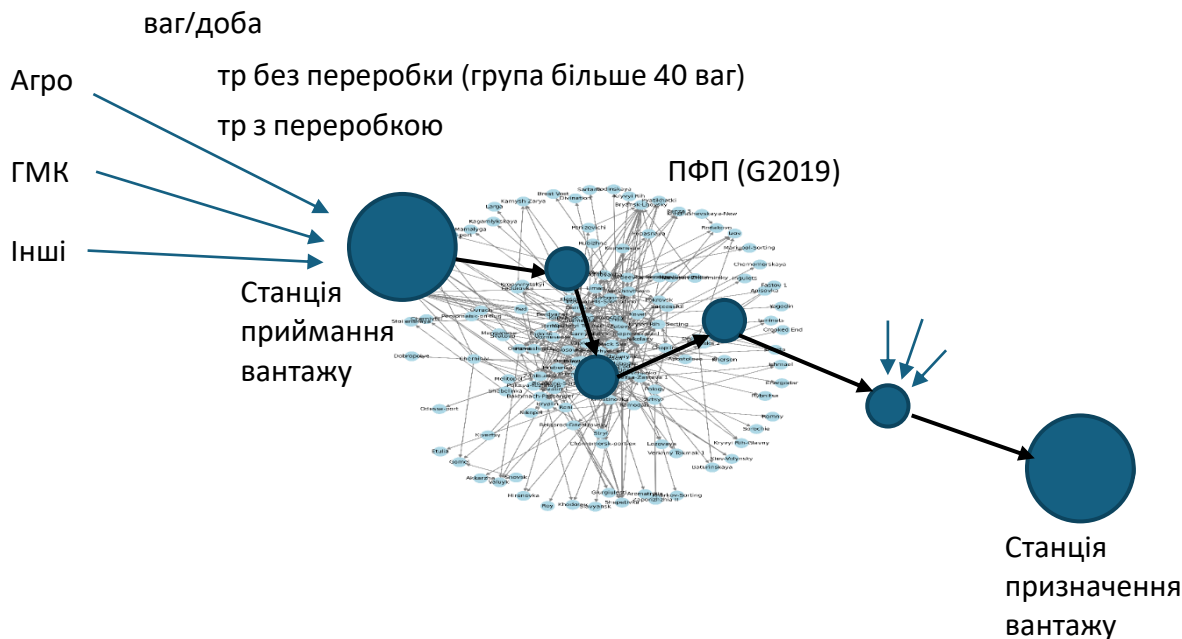


Рис. 4. Концептуальна схема інтеграції даних про навантаження вагонів із графом ПФП

Ураховуючи, що до ПФП входять лише опорні станції залізничної мережі (G2019 має 142 станції), а навантаження може відбуватися на більшій кількості станцій, що не входять у множину опорних, запропоновано, використовуючи вже згруповані дані файлу results.xlsx (рис. 3), здійснити за кожною станцією приймання вантажу для перевезення прив'язку до опорної, що входить до G2019. На рис. 5 наведена задача прив'язки станцій навантаження та призначення вантажу до опорних станцій ПФП. У роботі за допомогою спеціального алгоритму були визначені координати кожної станції відповідно до коду ЄСР. Це дало змогу в межах алгоритму, у випадку якщо станція приймання вантажу для перевезення не відповідала опорній станції ПФП, шукати найближчу станцію ПФП згідно з мінімальною відстанню. Для обчислення відстані між залізничними станціями, що задані їхніми географічними координатами  $(\varphi_1, \lambda_1)$  і  $(\varphi_2, \lambda_2)$ , використано формулу гаверсинусів

$$d = R \cdot c = R \cdot 2 \arctan 2(\sqrt{a}, \sqrt{1-a}), \quad (1)$$

$$a = \sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\Delta\lambda}{2}\right), \quad (2)$$

$$\Delta\varphi = \varphi_2 - \varphi_1, \quad \Delta\lambda = \lambda_2 - \lambda_1, \quad (3)$$

де  $\varphi_1, \lambda_1$  – широта і довгота станції приймання вантажу для перевезення або призначення вантажу, рад;

$\varphi_2, \lambda_2$  – широта і довгота опорної станції ПФП, рад;

$R$  – середній радіус Землі,  $R = 6371$  км;

$d$  – геодезична відстань між станціями.

Пошук відстаней за формулою гаверсинусів дає змогу коректно враховувати кривизну Землі, на відміну від евклідової метрики, що має похибки на великих відстанях або різних широтах. Після процедури прив'язки до станцій ПФП за методом алгоритму Дейкстри [17], що використовує функція `nx.shortest_path()` бібліотеки NetworkX [18], шукали маршрут до станції ПФП, що є останньою до станції призначення вантажу.

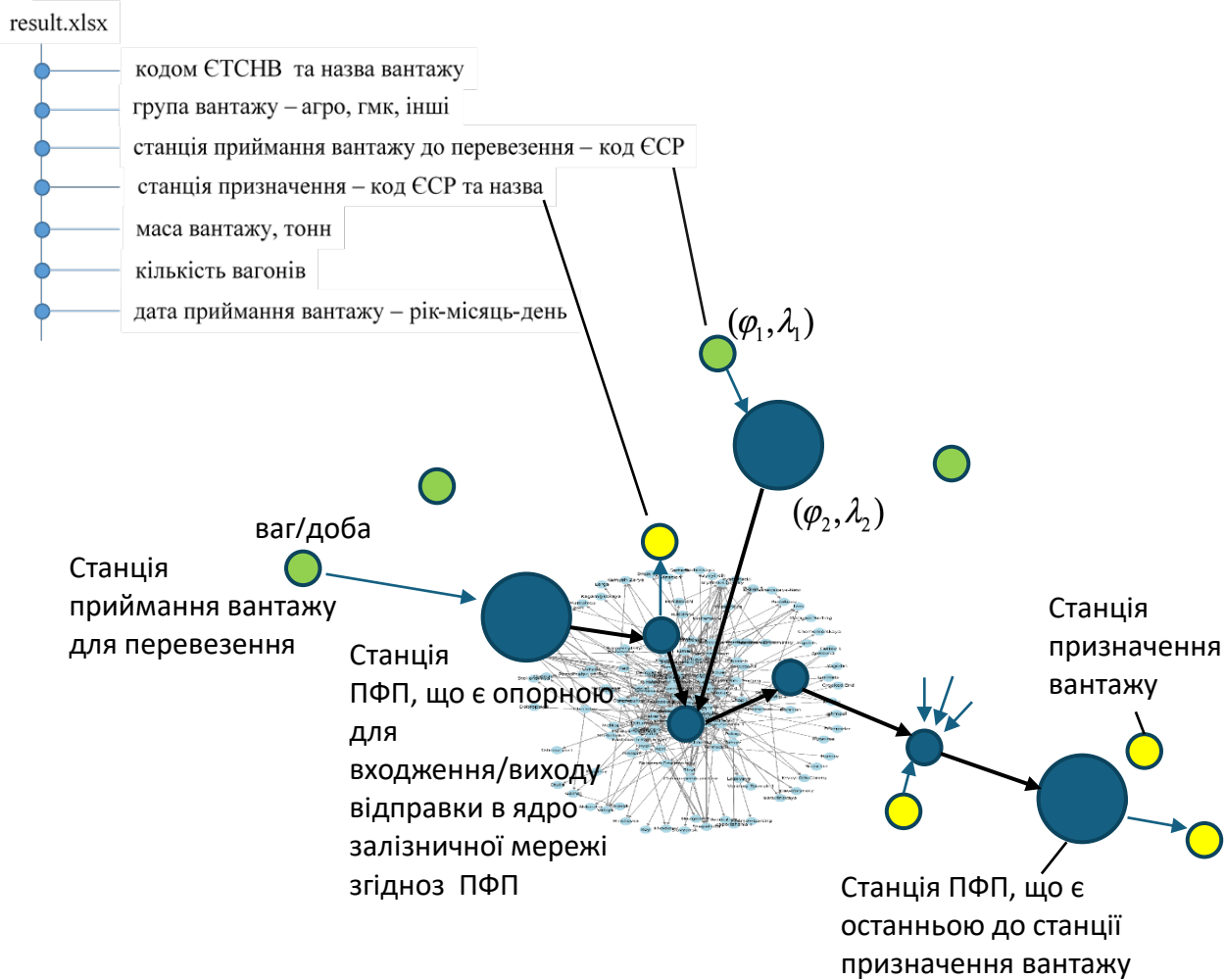


Рис. 5. Схема прив'язки станцій навантаження та призначення вантажу до опорних станцій ПФП

За результатами обчислень згенеровано файл routes.xlsx, у якому до вхідних даних із файлу results.xlsx додано три поля, що враховує послідовність назв станцій згідно зі знайденим маршрутом у графі ПФП, послідовність їхніх кодів і тип поїзда, що визначено порогом за кількістю вагонів (від 40 вагонів і більше – «маршрутна відправка» (М) відповідає типу вагонопотоку «транзит без переробки», інакше (П) – «транзит із переробкою»). На основі масиву даних файлу виконано процедуру постобробки результатів моделювання маршрутів для перетворення знайдених

послідовностей кодів станцій у відповідні номери рядків вихідного реєстру станцій all\_codes.txt. У межах процедури сформовано файл routes\_full.xlsx, у якому додано поле «Номери рядків», де перетворено послідовність кодів ЄСР станцій у відповідні номери рядків вихідного реєстру станцій файлу all\_codes.txt зі збереження характеристик станцій, що вона належить ПФП (якщо станція не належить ПФП, її номер від'ємний). Це необхідно для збереження інформації про тип станції (опорна, що належить ПФП, і допоміжна, що має навантаження) і коректного узгодження маршрутів із подальшими даними та моделями (рис. 6).

## routes\_full.xlsx

- код ЄТСНВ і назва вантажу
- група вантажу – агро, гмк, інші
- станція приймання вантажу для перевезення – код ЄСР
- станція призначення – код ЄСР і назва
- маса вантажу, т
- кількість вагонів
- дата приймання вантажу – рік-місяць-день
- найкоротший маршрут – послідовність назв станцій
- коди станцій маршрут – послідовність ЄСР станцій
- тип вагонопотоку – М чи П
- номер рядка – послідовність станцій закодованих номером рядка

Рис. 6. Структура масиву даних після процедури постобробки результатів моделювання маршрутів

Файл routes\_full.xlsx є вихідним для побудови навчальної матриці load.xlsx із масивом даних про навантаження розмірності (кількість станцій × кількість днів × вагонопотік), яка агрегує вагонопотік за маршрутом на кожній станції за днями і видами. Кожен день для станції зберігає рівно шість чисел у порядку, де закодовано поля для виду вантажу (агро, гмк, інші) і виду вагонопотоку (тр б/п, тр з/п) (рис. 7). Логіка процедури занесення відправки з відповідною

кількістю вагонів передбачає додавання кількості вагонів у поле станції, яка є початковою та в кожен наступну станцію маршруту, окрім останньої, що є станцією призначення цих вагонів. Послідовність станцій за маршрутом руху відправки вибирають у полі «номер рядка» файлу routes\_full.xlsx. За таким підходом відображено маршрут руху відправки згідно з ПФП, що дає змогу описати складні взаємозалежності впливу величини вагонопотоку та просторового його руху на затримки в мережі.

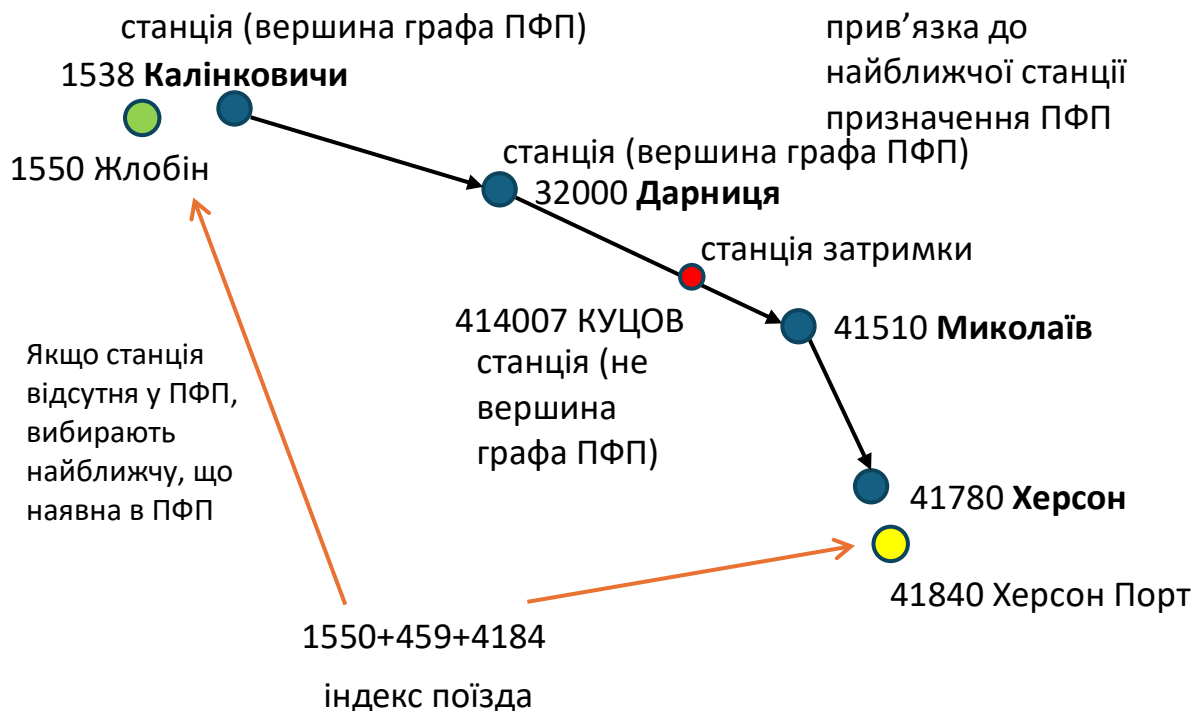
load.xlsx



Рис. 7. Структура масиву даних навчальної вибірки про вагонопотоки та схема кодування виду вагонопотоків для кожної станції ПФП

Для підготовки даних, що описують поїздо-години затриманих поїздів на станціях ПФП, розроблено процедуру на основі даних файлу datadelay.xlsx, де зазначені всі назви станцій затримки відповідних поїздів (рис. 2). Станції затримки часто є проміжними на дільницях, які не входять до опорних станцій ПФП. Важливо зменшити обчислювальний простір поставленої задачі. Із цією метою запропоновано інтегрувати дані станції з ПФП. Для цього ідентифіковано маршрут поїзного формування

за індексом поїзда, після чого в межах знайденої множини вершин запропоновано знаходити праву станцію з останньої пари сусідніх вузлів маршруту (призначення), між якими чисельно лежить код затримки. Наприклад, за рис. 8, станція затримки 41400 лежить у межах сусідньої пари кодів  $32000 < 41400 < 41510$ , і код станції затримки чисельно лежить між ними, чого достатньо для прив'язки до правого коду 41510.



Маршрут поїзда 15380 → 32000 → 41510 → 41780

Знайдено пару призначення 32000 → 41510

Чисельна перевірка, чи лежить код затримки  $32000 < 41400 < 41510$

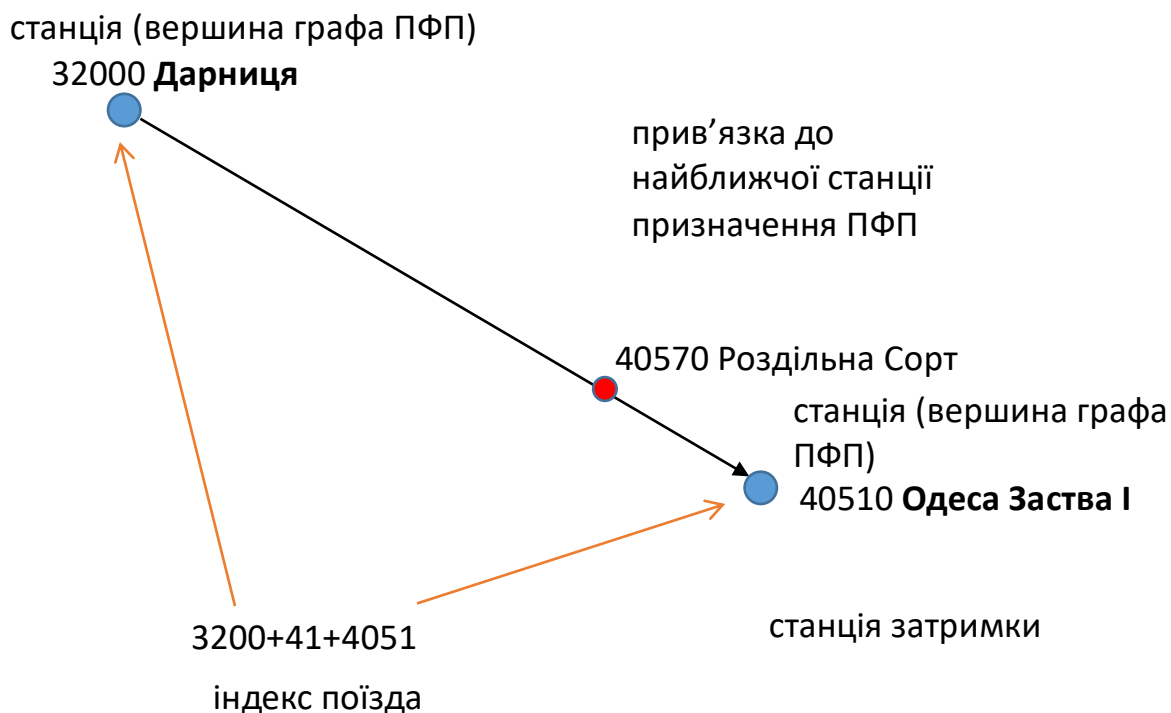
41510

прив'язка до правої станції призначення ПФП

Рис. 8. Приклад схеми прив'язки станції затримки, що не є станцією ПФП до опорної станції

За умови співпадіння маршруту руху поїзда з існуючим призначенням у ПФП, а код станції затримки не лежить на маршруті і не потрапляє «між парою» вузлів за числовим порядком, застосовують

принцип прив'язки до найближчого за значенням коду пари з пари сусідніх вузлів маршруту. Приклад схеми прив'язки станції затримки за правилом «найближча за кодом» наведено на рис. 9.



Маршрут поїзда 32000 → 40510

коду станції затримки немає серед вузлів маршруту 32000 41510

вибір за правилом «найближча за кодом»  $|40570 - 40510| = 60$

$|40570 - 32000| = 8570$

найменша різниця (60) з 40510

прив'язка до 40510

Рис. 9. Приклад схеми прив'язки станції затримки у випадку, якщо код станції затримки не лежить на маршруті і не потрапляє «між парою» вузлів за числовим порядком

За результатами інтеграції загальних даних про затримані поїзди на мережі до множини станцій ПФП згенеровано файл route\_report.xlsx, який є вихідним для побудови навчальної матриці для математичної моделі. Відповідно до співставлення рядків вхідної

матриці файлу load.xlsx згенеровано файл outputs.xlsx. У межах процедури перетворено тривалість простоїв у години по кожній даті та сформовано матрицю розмірністю 142 станції ПФП × дати, де в комірках зазначені поїздо-години простою. Структура масиву даних файлу outputs.xlsx наведена на рис. 10.

outputs.xlsx

	A	B	C	D	E
1	Код станції (5 цифр)	Назва станції	31.08.2019	01.09.2019	02.09.2019
2	13020	Брест-Вост.			
3	13940	Лунинец			
4	15000	Гомель			
5	15380	Калинковичи			
6	20000	Брянск-Льговский			
7	32000	Дарница		12,9	28,27694444
8	32050	Киев-Волынский			
9	32500	Чернигов			
10	32600	Бахмач-Пассажирский			

142 станції ПФП

12,9 поїздо-годин затримки

Рис. 10. Структура масиву даних навчальної вибірки про поїздо-години затримки

За результатами реалізації запропонованого підходу сформовано інтегрований набір даних, що поєднує операційні показники навантаження вагонів із графовою структурою плану формування поїздів. Розроблена процедура дала змогу виконати просторову прив'язку станцій навантаження та затримки поїздів до опорних станцій ПФП, а також сформувати маршрути руху вагонопотоків у межах мережі. У результаті було сформовано дві узгоджені матриці даних, зокрема вхідну матрицю load.xlsx, яка містить агреговані параметри навантаження за станціями ПФП, видами вантажів і типами вагонопотоків, і вихідну матрицю outputs.xlsx, що відображає поїздо-години затримок поїздів за станціями мережі.

Отриманий набір даних дає змогу інтегрувати просторову структуру залізничної мережі з операційними показниками перевізного процесу і створює інформаційну основу для побудови математичних моделей аналізу і прогнозування затримок поїздопотоків.

**Висновки.** У дослідженні розроблено метод інтеграції операційних даних функціонування залізничної мережі з графовою структурою плану формування поїздів, який формує інформаційну основу для аналізу та прогнозування затримок поїздопотоків у залізничній системі. У межах дослідження розроблено підхід щодо агрегування та перетворення операційних даних залізничної мережі, що включають інформацію про обсяги навантаження вагонів і випадки простою кинутих поїздів. Запропонований підхід допоміг узагальнити первинні дані за групами вантажів і видами вагонопотоків і перетворити їх у показники поїздо-годин затримок, які характеризують рівень порушення технології перевізного процесу на станціях мережі.

Сформовано метод інтеграції отриманих експлуатаційних показників із графовою моделлю плану формування поїздів, що дало змогу виконати

просторову прив'язку станцій навантаження та затримок до опорних станцій мережі і врахувати маршрути руху вагонопотоків у структурі ПФП. У результаті сформовано узгоджений набір даних, який поєднує операційні параметри перевізного процесу з просторовою структурою залізничної мережі та може бути використаний як інформаційна основа для подальшого аналізу і побудови математичних моделей прогнозування затримок поїздопотоків.

#### Список використаних джерел

1. Gurin D., Prokhorchenko A., Kravchenko M. & Shapoval G. Development of a method for modelling delay propagation in railway networks using epidemiological SIR models. *Eastern-European Journal of Enterprise Technologies*. 2020. Vol. 6 (3 (108)). P. 6–13. DOI: <https://doi.org/10.15587/1729-4061.2020.219285>
2. Official web-site of JSC Ukrzaliznytsia. URL: [https://www.uz.gov.ua/cargo\\_transportation](https://www.uz.gov.ua/cargo_transportation) (accessed: 20.02.2026) [in Ukrainian].
3. Lapamonpinyo P., Derrible S., Corman F. Real-time passenger train delay prediction using machine learning: A case study with Amtrak passenger train routes // *IEEE Open Journal of Intelligent Transportation Systems*. 2022. Vol. 3. P. 539–550. DOI: <https://doi.org/10.1109/OJITS.2022.3194879>
4. Wang P., Zhang Q. Train delay analysis and prediction based on big data fusion. *Transportation Safety and Environment*. 2019. Vol. 1 (1). P. 79-88. DOI: <https://doi.org/10.1093/tse/tdy001>
5. Huang P., Lessan J., Corman F., Meng L. Mining train delay propagation patterns from historical train operation data. Proceedings of the International Conference on Railway Operations Modelling and Analysis (RailNorrköping). *Linköping Electronic Conference Proceedings*.

## ІНФОРМАЦІЙНО-КЕРУЮЧІ СИСТЕМИ НА ЗАЛІЗНИЧНОМУ ТРАНСПОРТІ

2019. Vol. 69:29. P. 439-451. URL: <https://ep.liu.se/ecp/069/029/ecp19069029.pdf>
6. Fu Q., Ding S., Zhang T., Wang R., Hu P., Pu C. Short-term train arrival delay prediction: a data-driven approach. *Railway Sciences*. 2024. Vol. 3, No. 4. P. 514–529. DOI: <https://doi.org/10.1108/rs-04-2024-0012>
  7. Zhang D., Peng Y., Zhang Y., Wu D., Wang H., Zhang H. Train Time Delay Prediction for High-Speed Train Dispatching Based on Spatio-Temporal Graph Convolutional Network. *IEEE Transactions on Intelligent Transportation Systems*. 2022. Vol. 23, No. 3. P. 2434–2444. DOI: <https://doi.org/10.1109/TITS.2021.3093840>
  8. Li J., Xu X., Wang J. A Train Delay Prediction Approach for High-Speed Railway Considering Spatiotemporal Logical Relationships in Train Operation. *SSRN Electronic Journal*. 2024. DOI: <https://doi.org/10.2139/ssrn.4828128>
  9. Пархоменко Л. О., Прохоров В. М. Уточнення моделі розрахунку плану формування поїздів за рахунок урахування ефекту навантаженості сортувальних станцій. *Вісник Приазовського державного технічного університету. Серія «Технічні науки»*. 2024. № 48. С. 193–205. DOI: <https://doi.org/10.31498/2225-6733.48.2024.310712>
  10. Butko T., Prymachenko H., Kovalov A., Tarasov K., Kolisnyk A. Research on the Issue of Prognosticationing the Volume of Passenger Traffic on Railway Transport in Modern Conditions. *Review of Economics and Finance*. 2023. Vol. 21. P. 236–245. DOI: <https://doi.org/10.55365/1923.x2023.21.22>
  11. Dmytro D. Economic efficiency of railway in Ukraine: railway economies of scale, scope and density : магістерська робота. Київ : Kyiv School of Economics, 2022. URL: [https://kse.ua/wp-content/uploads/2023/07/Dushko-Dmytro\\_.pdf](https://kse.ua/wp-content/uploads/2023/07/Dushko-Dmytro_.pdf) (дата звернення: 13.03.2026).
  12. Габа В., Стрелько О., Грушевська Т., Гусар І. Дослідження залізничних перевезень вантажів груповими та одиночними вагонами. *Транспортні системи і технології*. 2021. Вип. 36. С. 150–162. DOI: <https://doi.org/10.32703/2617-9040-2020-36-16>
  13. Li Z., Huang P., Wen C., Dong W., Ji Y., Rodrigues F. Railway network delay evolution: A heterogeneous graph neural network approach. *Applied Soft Computing*. 2024. Vol. 159. Article 111640. DOI: <https://doi.org/10.1016/j.asoc.2024.111640>
  14. Peng T., Gan M., Ou Q., Yang X., Wei L., Ler H. R., Yu H. Railway cold chain freight demand forecasting with graph neural networks: A novel GraphARMA-GRU model. *Expert Systems with Applications*. 2024. Vol. 255, Part B. Article 124693. DOI: <https://doi.org/10.1016/j.eswa.2024.124693>
  15. Єдина тарифно-статистична номенклатура вантажів. Офіційний сайт АТ Укрзалізниця. URL: [https://uz.gov.ua/cargo\\_transportation/legal\\_documents/nomenklatura/changes\\_etcnb/](https://uz.gov.ua/cargo_transportation/legal_documents/nomenklatura/changes_etcnb/) (дата звернення 27.01.2026).
  16. Правила приймання вантажів до перевезення: Наказ Міністерства транспорту України. Про затвердження окремих розділів Правил перевезення вантажів : Наказ Міністерства транспорту України від 21 листопада 2000 р. № 644; зареєстр. в Міністерстві юстиції України 24 листопада 2000 р. за № 861/5082. База даних «Законодавство України» / Верховна Рада України. URL: <https://zakon.rada.gov.ua/laws/show/z0861-00#n49> (дата звернення: 28.10.2025).
  17. Dijkstra E. A note on two problems in connexion with graphs. *Numer. Math.* 1959. 1. 269–271.
  18. NetworkX Developers. Shortest Paths – NetworkX Documentation. URL: [https://networkx.org/documentation/stable/reference/algorithms/shortest\\_paths.html](https://networkx.org/documentation/stable/reference/algorithms/shortest_paths.html) (дата звернення: 15.01.2026).

## References

1. Gurin, D., Prokhorchenko, A., Kravchenko, M., & Shapoval, G. (2020). Development of a method for modelling delay propagation in railway networks using epidemiological SIR models. *Eastern-European Journal of Enterprise Technologies*, 6(3(108)), 6–13. <https://doi.org/10.15587/1729-4061.2020.219285>
2. JSC Ukrzaliznytsia. (n.d.). *Cargo transportation*. Retrieved February 20, 2026, from [https://www.uz.gov.ua/cargo\\_transportation](https://www.uz.gov.ua/cargo_transportation)
3. Lapamonpinyo, P., Derrible, S., & Corman, F. (2022). Real-time passenger train delay prediction using machine learning: A case study with Amtrak passenger train routes. *IEEE Open Journal of Intelligent Transportation Systems*, 3, 539–550. <https://doi.org/10.1109/OJITS.2022.3194879>
4. Wang, P., & Zhang, Q. (2019). Train delay analysis and prediction based on big data fusion. *Transportation Safety and Environment*, 1(1), 79–88. <https://doi.org/10.1093/tse/tdy001>
5. Huang, P., Lessan, J., Corman, F., & Meng, L. (2019). Mining train delay propagation patterns from historical train operation data. *Linköping Electronic Conference Proceedings*, 69(29), 439–451. <https://ep.liu.se/ecp/069/029/ecp19069029.pdf>
6. Fu, Q., Ding, S., Zhang, T., Wang, R., Hu, P., & Pu, C. (2024). Short-term train arrival delay prediction: A data-driven approach. *Railway*

- Sciences*, 3(4), 514–529. <https://doi.org/10.1108/rs-04-2024-0012>
7. Zhang, D., Peng, Y., Zhang, Y., Wu, D., Wang, H., & Zhang, H. (2022). Train time delay prediction for high-speed train dispatching based on spatio-temporal graph convolutional network. *IEEE Transactions on Intelligent Transportation Systems*, 23(3), 2434–2444. <https://doi.org/10.1109/TITS.2021.3093840>
  8. Li, J., Xu, X., & Wang, J. (2024). A train delay prediction approach for high-speed railway considering spatiotemporal logical relationships in train operation. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4828128>
  9. Parkhomenko, L. O., & Prokhorov, V. M. (2024). Utochnennia modeli rozrakhunku planu formuvannia poizdiv za rakhunok urakhuvannia efektu navantazhenosti sortovalnykh stantsii [Refinement of the model for calculating the train formation plan by taking into account the effect of marshalling yards loading]. *Visnyk Pryazovskoho derzhavnoho tekhnichnoho universytetu. Seriya «Tekhnichni nauky»* [Reporter of the Priazovskiy State Technical University. Section: Technical Sciences], (48), 193–205. <https://doi.org/10.31498/2225-6733.48.2024.310712>
  10. Butko, T., Prymachenko, H., Kovalov, A., Tarasov, K., & Kolisnyk, A. (2023). Research on the issue of prognostication of the volume of passenger traffic on railway transport in modern conditions. *Review of Economics and Finance*, 21, 236–245. <https://doi.org/10.55365/1923.x2023.21.22>
  11. Dushko, D. (2022). *Economic efficiency of railway in Ukraine: Railway economies of scale, scope and density* [Master's thesis, Kyiv School of Economics]. [https://kse.ua/wp-content/uploads/2023/07/Dushko-Dmytro\\_.pdf](https://kse.ua/wp-content/uploads/2023/07/Dushko-Dmytro_.pdf)
  12. Haba, V., Strelko, O., Hrushevska, T., & Husar, I. (2021). Doslidzhennia zaliznychnykh perevezhen vantazhiv hrupovymy ta odynochnymy vahonamy [Study of railway freight transport by group and single wagons]. *Transportni systemy i tekhnolohii* [Transport Systems and Technologies], (36), 150–162. <https://doi.org/10.32703/2617-9040-2020-36-16>
  13. Li, Z., Huang, P., Wen, C., Dong, W., Ji, Y., & Rodrigues, F. (2024). Railway network delay evolution: A heterogeneous graph neural network approach. *Applied Soft Computing*, 159, Article 111640. <https://doi.org/10.1016/j.asoc.2024.111640>
  14. Peng, T., Gan, M., Ou, Q., Yang, X., Wei, L., Ler, H. R., & Yu, H. (2024). Railway cold chain freight demand forecasting with graph neural networks: A novel GraphARMA-GRU model. *Expert Systems with Applications*, 255(Part B), Article 124693. <https://doi.org/10.1016/j.eswa.2024.124693>
  15. JSC Ukrzaliznytsia. (n.d.). *Iedyna taryfno-statystychna nomenklatura vantazhiv* [Unified tariff and statistical nomenclature of goods]. Retrieved January 27, 2026, from [https://uz.gov.ua/cargo\\_transportation/legal\\_documents/nomenklatura/changes\\_etcnb/](https://uz.gov.ua/cargo_transportation/legal_documents/nomenklatura/changes_etcnb/)
  16. Verkhovna Rada of Ukraine. (2000). *Pravylyn pryimannia vantazhiv do perevezennia: Nakaz Ministerstva transportu Ukrainy No. 644* [Rules for accepting goods for transportation: Order of the Ministry of Transport of Ukraine No. 644]. Baza danykh «Zakonodavstvo Ukrainy». Retrieved October 28, 2025, from <https://zakon.rada.gov.ua/laws/show/z0861-00#n49>
  17. Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1, 269–271.
  18. NetworkX Developers. (n.d.). *Shortest paths*. NetworkX Documentation. Retrieved January 15, 2026, from [https://networkx.org/documentation/stable/reference/algorithms/shortest\\_paths.html](https://networkx.org/documentation/stable/reference/algorithms/shortest_paths.html)

KYMAN A. M., PROKHORCHENKO H. O., Kravchenko M. A.

**Development of a method for integrating railway operational data with the graph structure of the train formation plan**

**Abstract.** *The paper proposes a method for integrating operational railway network data with the graph structure of the train formation plan. The method is aimed at creating an information basis for the analysis and forecasting of train flow delays while taking into account the spatial structure of the railway transport system. The study uses operational data of JSC «Ukrzaliznytsia» related to wagon loading of different freight categories as well as information on idle «abandoned» trains at stations within the operational area of the Odesa Railway regional branch. A data preprocessing procedure is developed that aggregates operational records according to freight groups and wagon flow types. The obtained data are further transformed into indicators of train-hours of delay, which characterize disruptions in the technological process of railway transportation. To incorporate the spatial structure of the railway network, a method is proposed for integrating loading data and delay information with the graph representation of the train formation plan. The proposed approach includes an algorithm for assigning loading stations and delay stations to the main stations of the train formation network using geographical coordinates and graph-based routing procedures. This makes it possible to reconstruct wagon flow routes within the network and to represent the interaction between freight flows and railway infrastructure nodes. As a result, an integrated dataset was generated that combines operational indicators of wagon loading, spatial routes of freight flows, and delay indicators expressed in train-*

hours. The resulting data structure forms a consistent input and output dataset that can be used as a training sample for the development of mathematical models for the analysis and prediction of train delays in railway networks. The proposed method enables the integration of operational transportation data with the spatial topology of the railway network and can serve as a basis for further research on delay propagation and predictive modeling of railway transport processes.

**Keywords:** railway network, train formation plan, graph model, operational transportation data, wagon flows, train delays, delay prediction.

Стаття поширюється на умовах ліцензії Creative Commons Attribution License International CC-BY.

Киман Андрій Миколайович, кандидат технічних наук, доцент, доцент кафедри управління вантажною і комерційною роботою, Український державний університет залізничного транспорту.

ORCID: <http://orcid.org/0000-0002-4000-3287>.

E-mail: [kyman.am@kart.edu.ua](mailto:kyman.am@kart.edu.ua).

Прохорченко Галина Олегівна, кандидат технічних наук, доцент кафедри управління експлуатаційною роботою, Український державний університет залізничного транспорту.

ORCID iD: <https://orcid.org/0000-0003-1158-3953>.

E-mail: [g.o.prokhorchenko@kart.edu.ua](mailto:g.o.prokhorchenko@kart.edu.ua).

Кравченко Михайло Анатолійович, кандидат технічних наук, докторант кафедри управління експлуатаційною роботою, Український державний університет залізничного транспорту.

ORCID iD: <https://orcid.org/0000-0001-7445-8952>.

E-mail: [kravchenko.ma@kart.edu.ua](mailto:kravchenko.ma@kart.edu.ua).

Kyman Andrii, PhD (Tech), Associate Professor, Associate Professor of the Department of Freight and Commercial Operations Management, Ukrainian State University of Railway Transport.

ORCID: <http://orcid.org/0000-0002-4000-3287>.

E-mail: [kyman.am@kart.edu.ua](mailto:kyman.am@kart.edu.ua).

Prokhorchenko Halyna, PhD (Tech), Associate Professor of the Department of Operations Management, Ukrainian State University of Railway Transport.

ORCID iD: <https://orcid.org/0000-0003-1158-3953>.

E-mail: [g.o.prokhorchenko@kart.edu.ua](mailto:g.o.prokhorchenko@kart.edu.ua).

Kravchenko Mykhailo, PhD (Tech), doctoral student of the Department of Operations Management, Ukrainian State University of Railway Transport.

ORCID iD: <https://orcid.org/0000-0001-7445-8952>.

E-mail: [kravchenko.ma@kart.edu.ua](mailto:kravchenko.ma@kart.edu.ua).

Стаття надійшла 02.03.26

Стаття прийнята до друку після рецензування 10.04.26

Стаття опублікована (оприлюднена) 29.05.26