Tunnelling and Underground Space Technology 174 (2026) 107708

Contents lists available at ScienceDirect

Tunnelling and Underground Space Technology incorporating Trenchless Technology Research

journal homepage: www.elsevier.com/locate/tust

Analysis of drainage system defects using co-occurrence patterns

Dramani Arimiyaw [a], Tarek Zayed [a], Jingchao Yang [a,*], Beenish Bakhtawar [a], Sherif Abdelkhalek [a], Mohamed Nashat [a,b]

[a] Department of Building and Real Estate (BRE), Faculty of Construction and Environment (FCE), The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

[b] Department of Public Works Engineering, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt

ARTICLE INFO

Keywords:
Sewer deterioration mechanisms
Defect co-occurrence
Network analysis
Louvain clustering
Infrastructure asset management
Proactive maintenance

ABSTRACT

Urban drainage systems face unprecedented challenges from ageing infrastructure, environmental stressors, and increasing service demands. Traditional condition assessment approaches aggregate diverse defect patterns into single condition scores, obscuring the underlying deterioration mechanisms that drive failure processes. This study develops a network science framework to identify systematic defect co-occurrence patterns using CCTV inspection data from Hong Kong's drainage network. Constructing weighted defect co-occurrence networks and applying Louvain clustering algorithms, we identified four distinct deterioration mechanisms: Lining-Deformation Co-occurrence Pattern (18.4% prevalence), Structural-Hydraulic Co-occurrence Pattern (54.5%), Root-Joint Co-occurrence Pattern (22.3%), and Connection-Sediment Co-occurrence Pattern (4.9%). Multi-method validation using five alternative weighting schemes demonstrated robust clustering consistency (Adjusted Rand Index = 0.438–1.000), while statistical analysis revealed significant associations between mechanisms and infrastructure characteristics (diameter, age, material) and contextual factors (district, land use, traffic intensity). These findings support the development of mechanism-informed management strategies, enabling utilities to move from purely reactive condition-based approaches toward targeted interventions informed by systematic defect association patterns. While cross-sectional analysis cannot establish causal relationships, the observed co-occurrence patterns provide actionable intelligence for risk-based asset management.

1. Introduction

Urban drainage infrastructure represents one of the most critical yet vulnerable components of modern water management systems, with networks worldwide confronting unprecedented challenges in maintaining structural integrity and operational performance (Carvalho et al., 2018; Fontecha et al., 2021; Yang et al., 2025). These systems safely transport both foul water and urban runoff separately through the sewer and stormwater pipe systems, respectively, to treatment plants and water bodies, avoiding contamination and flooding (Harvey & McBean, 2014a; Nguyen et al., 2022). In developed countries, prolonged usage (between 50 and 75 years) and external influences such as climate change, ground movement, and other factors cause the existing drainage system to suffer significant damage, leading to defects (Harvey and McBean, 2014a; Loganathan et al., 2024; Yang et al., 2025). An estimated €28-32 billion investment is required for Norway's water infrastructure, $14 billion replacement cost per year in the US, and $109 billion for the replacement of 35% of wastewater and 23% of stormwater infrastructure in Canada (Li et al., 2019; Nguyen et al., 2022; Woldesellasse & Tesfamariam, 2024; Yin et al., 2020). In Hong Kong, 22 km of stormwater drains and sewers were rehabilitated in 2016–2017 at a cost of HK$138 million (Drainage Services Department, 2017). Pipe defects, such as cracks, holes, and fractures, that form in ageing pipes become discharge points for untreated sewage into the environment and entry points for runoff water and tree roots (Arimiyaw et al., 2025; Harvey & McBean, 2014a). These sewage-driven contaminants can reach depths of over 60m when cracks and fissures are present in the subsurface (Harvey & McBean, 2014b). Therefore, regular inspection and diagnosis of the drainage infrastructure are paramount to ensure its sustainability and prevent severe consequences, especially in wastewater systems that contain toxic chemicals.

Condition assessment is a key component of asset management,

* Corresponding author.
E-mail address: jingchao-aaron.yang@connect.polyu.hk (J. Yang).

https://doi.org/10.1016/j.tust.2026.107708
Received 5 December 2025; Received in revised form 20 March 2026; Accepted 18 April 2026
Available online 27 April 2026

providing updated information on the current and future states of infrastructure (Yang et al., 2025a; Yang et al., 2025). In recent years, CCTV-enabled condition assessment has been used to estimate the remaining service life of drainage pipes (Kumar et al., 2024; Sousa et al., 2014). In this process, individual defects are assigned scores based on physical grading protocols and then aggregated at the segment level as a condition grade using peak, mean, and total scores or calculation methods (Elmasry et al., 2017; Khazraeializadeh et al., 2014). Pipes are usually assigned grades 1–5, with 1 indicating excellent and 5 indicating the worst condition. Nonetheless, the current condition assessment approach faces several critical challenges. The process is primarily reactive, as inspections are conducted in response to failures that cause financial losses and environmental concerns (Nashat et al., 2025). Accessibility and resource constraints have limited network inspections to approximately 10–23% (Abdelkhalek & Zayed, 2023; Balekelayi & Tesfamariam, 2020). While CCTV provides detailed documentation of pipe defects through standardised coding systems, utilities face a disconnect between their data-collection capabilities and the mechanistic insights that can be derived from them.

Alternatively, predictive modeling as a proactive strategy has gained significant attention (Yang et al., 2025; Zamanian and Shafieezadeh, 2023). This strategy is supported by modeling techniques that extract information from existing inspection datasets to predict pipe condition for scheduled maintenance (Harvey & McBean, 2014a; Nguyen & Seidu, 2022). Statistical and machine learning models have been dominantly used in drainage predictive maintenance. For instance, Chughtai and Zayed (2008) developed multiple regression models to predict sewer conditions using historical physical, operational, and environmental data from two cities in Canada. The developed models achieved 82–86% accuracy on the validation set. Lubini and Fuamba employed a multinomial logistic regression model to predict the deterioration state of sewer conduit in Quebec (Lubini & Fuamba, 2011). They achieved 54% accuracy for pipes in excellent condition (grade 1–2) and 59% for fair condition (grade 3). However, these models are limited in their ability to handle complex predictive factors. They assume linear interactions among factors, which do not accurately reflect the drainage environment (Elmasry et al., 2017). On the other hand, advanced machine learning tools, including support vector machines, artificial neural networks, decision trees, and random forests, are currently used to assess drainage pipe conditions. These models have offered many advantages over statistical models. They are able to handle the growing availability of inspection data and the complex interaction of the factors (Jingchao Yang et al., 2025). Additionally, they demonstrate superior pattern recognition capabilities, enabling them to accurately capture the underlying deterioration pattern. Furthermore, these models incorporate feature importance metrics that quantify each feature's contribution, enabling a more targeted utility intervention.

However, despite advanced machine learning capabilities for accurate condition prediction, these methods pose critical challenges. These predictions focus exclusively on condition grades, typically ordinal scales (1–5) or binary failures (0/1), reducing complex multi-defect patterns into single numeric scores (Abdelkhalek & Zayed, 2023; Chughtai & Zayed, 2008; Harvey & McBean, 2014b; Mashford et al., 2011; Nguyen & Razak, 2023). A pipe experiencing joint displacement, longitudinal cracking, and infiltration receives the same grade (e.g., 4) classification as one with surface damage, root intrusion, and sediment deposits, based on severity, despite these representing fundamentally different physical processes that require distinct intervention strategies. This aggregation conceals critical information about the failure mechanisms embedded in the defect co-occurrence pattern. When multiple defects consistently appear together across multiple pipe segments, they likely reflect a common underlying deterioration mechanism.

Recognising these limitations, recent studies have emphasised the importance of a defect-based framework for prioritising drainage maintenance. In an expert-based survey, Alqahtani et al. studied common structural defects in sewer pipelines and identified their causative relationship (Alqahtani et al., 2023). It is demonstrated that some defects are accelerated by the presence of other defects in the pipes. For instance, steep slopes with high velocity are likely to generate hydrogen sulfide, which can cause corrosion and lead to cracks and fractures. Also, groundwater infiltration through breaks and cracks can carry sediments or soil particles into the pipe, potentially causing blockage. Similarly, Daher et al. evaluated the relative weights of network components and general defect types, such as structural, operational, and installation defects, using the Analytical Network Process (ANP) as an aggregation method (Daher et al., 2018). While these studies provide valuable expert-driven insights, they reveal four critical knowledge gaps in current literature. First, no study has empirically validated whether drainage pipe defects co-occur in statistically non-random patterns using large-scale CCTV inspection databases; existing predictive work either treats defects as independent predictors or aggregates them into single condition scores. Second, while network science has proven effective for pattern discovery in other complex systems (such as symptom co-occurrence networks in healthcare or species interactions in ecology), it has not yet been applied to drainage deterioration analysis. Third, early warning capability remains undeveloped: certain defects may serve as central indicators signalling potential cascading failures, but without systematic analysis of co-occurrence patterns, these critical signals go unrecognised until severe structural damage occurs. Fourth, no research translates defect relationship patterns into actionable maintenance protocols that move beyond aggregate condition assessment.

This study develops and validates a network science framework by applying cluster detection methods to CCTV inspection data from Hong Kong's drainage network. Specifically, we construct defect co-occurrence networks and apply the Louvain clustering algorithm to identify clusters that represent potential deterioration mechanisms. This framework addresses four key challenges:

1. Distinguishing true systematic co-occurrence that reflects physical mechanisms from spurious patterns arising from sampling bias or data artefacts. By constructing defect co-occurrence networks and identifying clusters, we test whether drainage pipe deterioration follows discoverable mechanisms rather than random processes, unlike existing approaches that treat defects independently.
2. Identifying which defects function as structurally central nodes within each mechanism that may serve as candidate early-warning indicators, enabling potential predictive intervention before catastrophic failure. Centrality analysis is conducted to examine the degree of association between defects within the cluster and the specific implications of defect positions.
3. Characterising identified mechanisms in terms of defect composition, prevalence, structural relationships (rather than strictly temporal progression), and associations with pipe and environmental attributes. Through comprehensive validation that combines parameter sensitivity analysis and comparisons of alternative metrics, we determine whether the identified patterns represent genuine deterioration mechanisms or methodological artefacts.
4. Integrating mechanism knowledge into maintenance protocols that move beyond condition scores to capture how and why pipes deteriorate. The resulting mechanism-based framework enables transformation from aggregate condition prediction to mechanistic deterioration management, supporting proactive asset management strategies such as risk-stratified inspection prioritisation and mechanism-specific rehabilitation.

The rest of the paper is structured and organised as follows: Section 2 describes the study context, including Hong Kong's drainage network characteristics, CCTV inspection protocols, and the dataset comprising pipe segments with documented defects. Section 3 presents the analytical methodology, detailing the construction of the defect co-occurrence network, cluster detection algorithms, and validation frameworks.

Section 4 reports results in four parts: defect co-occurrence patterns, identified deterioration mechanisms, validation outcomes demonstrating pattern robustness, and associations between mechanisms and infrastructure attributes. Section 5 discusses theoretical implications for understanding deterioration processes and practical applications. Section 6 concludes by synthesizing key findings, articulating contributions, and outlining future research directions.

2. Study area and data

Hong Kong, situated on the southeastern coast of China, is used as a case study. It is a highly urbanised city with a high population density: about 7.5 million people in a land area of 1,100 km^2. Hong Kong experiences heavy rainfall during the rainy season, with average annual rainfall reaching 2400 mm (Chui et al., 2006; Keung et al., 2018), putting pressure on its drainage infrastructure. The Drainage Services Department (DSD) is responsible for managing the drainage infrastructure, including both stormwater and sewage systems. Hong Kong has adopted a separate drainage system, isolating both infrastructures; they currently manage about 2400 km of stormwater drains and 1900 km of sewer pipes. DSD has reported that about 1800 km (over 30%) have been in service for 30 years or more, with many of them showing signs of wear and tear. Therefore, regular inspections are conducted using CCTV cameras to ascertain the condition of the drainage system.

Fig. 1 illustrates the comprehensive data collection and integration workflow employed in this study. The analysis utilised an extensive repository comprising CCTV inspection reports from 5,470 folders containing 30,241 files from the DSD, collected over a 14-year operational period (2007–2021). This comprehensive dataset, totalling 75.8 GB, provides extensive spatial and temporal coverage of the urban drainage network, ensuring a representative sampling of diverse operational conditions and deterioration stages across Hong Kong's drainage infrastructure (Jingchao Yang et al., 2025). To enable comprehensive defect clustering analysis and subsequent infrastructure-defect association studies, the core defect data were systematically integrated with complementary infrastructure and environmental information from multiple Hong Kong government agencies. Building on primary defect identification from CCTV inspection records, the study incorporated comprehensive pipeline attributes from DSD, including district, length, material, age, and diameter, to characterise infrastructure conditions. Additionally, the dataset was further enriched by integrating key environmental variables, including Annual Average Daily Traffic (AADT) data from the Transport Department and land-use classifications from the Planning and Lands Departments, along with other relevant spatial and demographic variables.

Data preprocessing addressed three critical challenges: data quality assurance, format standardisation, and defect record interpretation. The original CCTV inspection data embedded in PDF reports required systematic extraction and conversion processes. Raw PDF files were first reorganised from hierarchical folder structures and processed using Python libraries (PyPDF2, PDFMiner, and pdfrw) for keyword-based clustering. The PDF reports varied in format and were processed using specialised extraction software to convert tabular data into a structured Excel format. This multi-stage extraction process handled variations in layout structures, overlapping text elements, and inconsistent formatting across different reporting periods. A systematic multi-stage preprocessing pipeline was implemented to ensure data quality and analytical robustness. Following the PDF-to-Excel conversion, extracted data underwent comprehensive validation and standardisation procedures. The analysis focused on pipe segments exhibiting multiple co-occurring defects, with an initial subset identified through defect field analysis. Logical inconsistencies were removed, duplicate records were retained only for the most complete records, and segments with missing geolocation data were excluded. The analysis was constrained to concrete and vitrified clay pipes, the two predominant materials in the network, and missing values were imputed using appropriate statistical

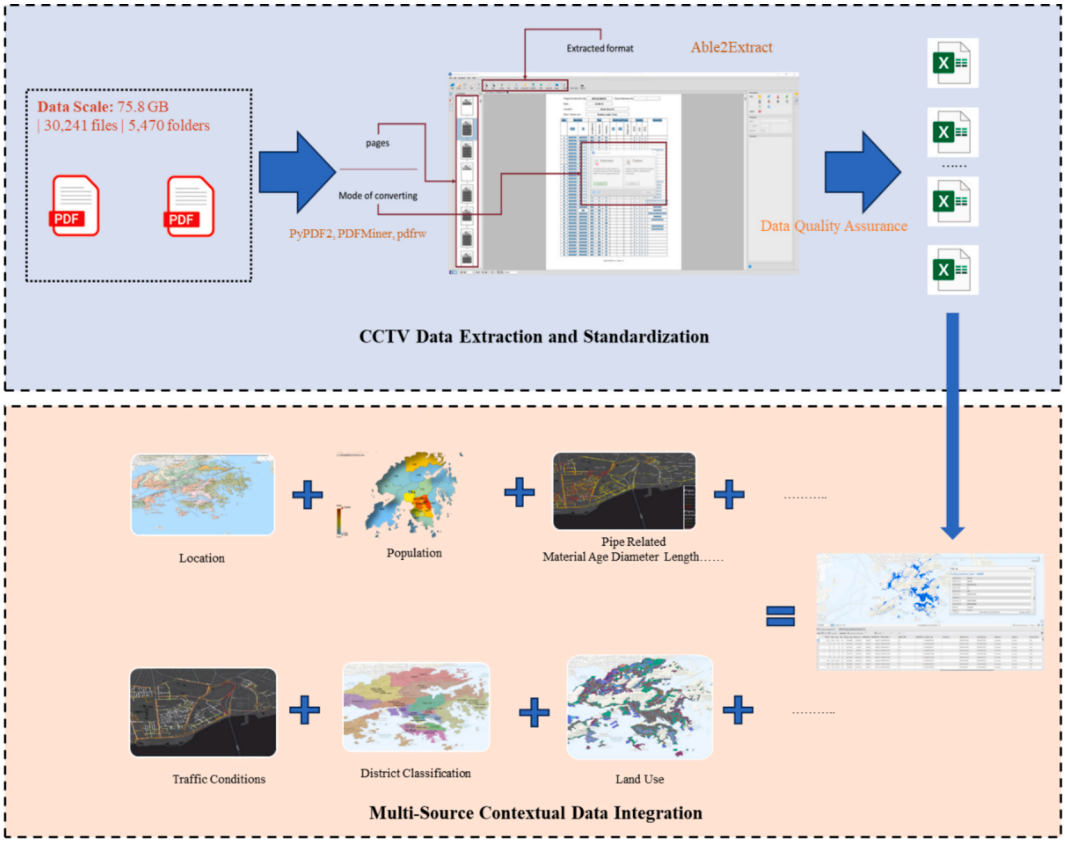

Fig. 1. Comprehensive Data Collection and Integration Workflow.

methods (categorical variables: mode; continuous variables: median) to preserve sample size without introducing significant bias. All datasets undergo regular updates and stringent quality control measures by their respective agencies to ensure accuracy and consistency across different data sources. This multi-agency collaboration establishes a robust foundation for comprehensive pipeline condition analysis.

3. Methodology

Fig. 2 outlines the four-stage analytical framework employed to investigate systematic deterioration patterns in drainage infrastructure, encompassing data processing, network clustering, infrastructure association analysis, and maintenance strategy development.

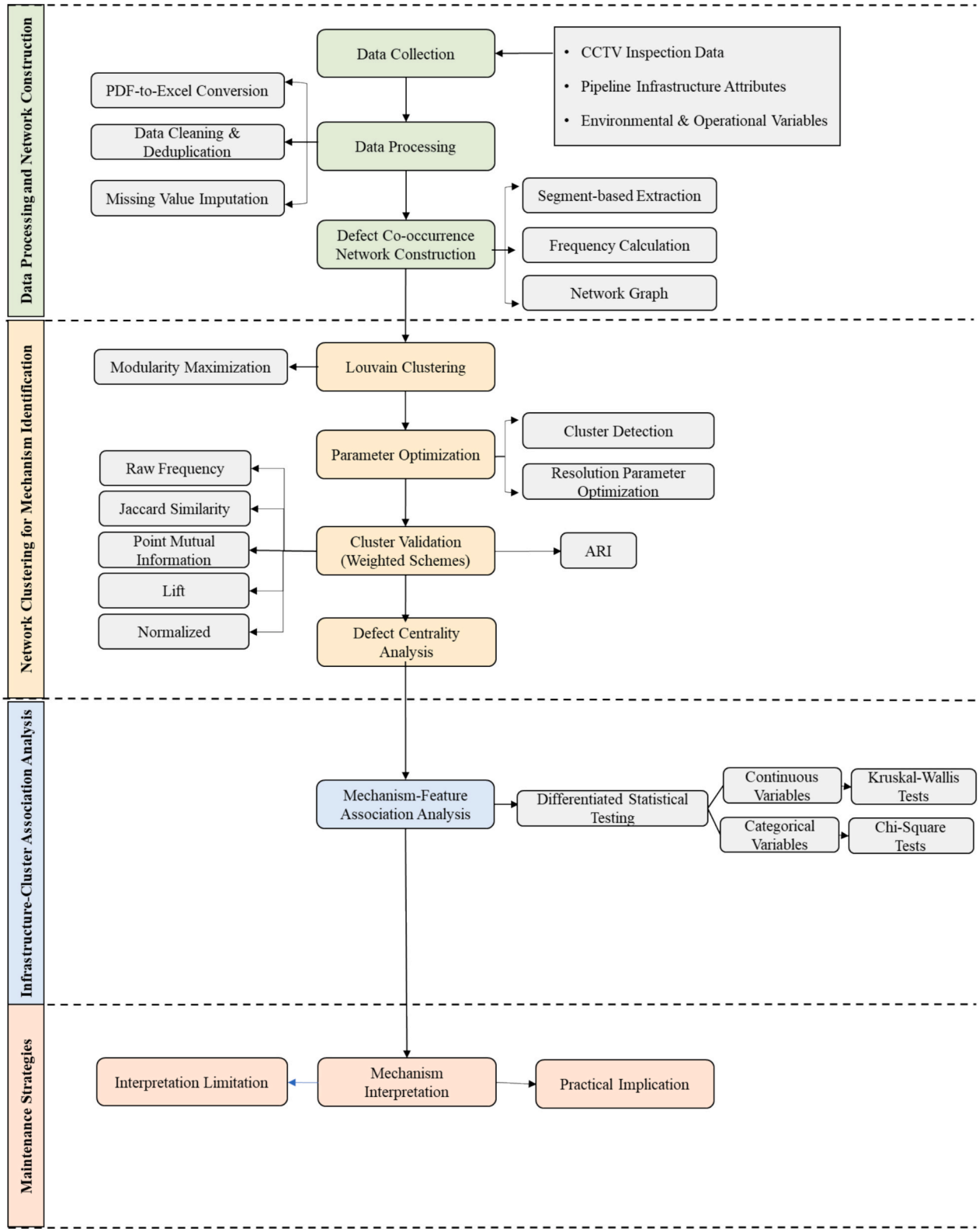

Fig. 2. Research methodology.

3.1. Defect co-occurrence network construction

This study examined drainage defects using a framework in which co-occurrence is defined as the presence of multiple defects documented within the same pipe segment between consecutive manholes during a single CCTV inspection. This approach recognises that defects appearing together in the same segment may share common underlying deterioration drivers or represent different manifestations of the same physical mechanism. It is important to note that this segment-level co-occurrence definition does not capture the temporal sequence of defect emergence; defects observed together during a single inspection may have emerged at different times. This limitation and its implications for causal interpretation are discussed in Section 5. Each pipe segment p is characterised by its defect profile $D_p = \{d_1, d_2, \cdots..., d_k\}$ where $D = \{d_1, d_2, d_3, \cdots..., d_m\}$, represents the set of all distinct defect types observed in the network. A defect d_i is present in the segment p if $d_i \in D_p$. The co-occurrence frequency between any defect pair (d_i, d_j) quantifies how often these defects appear together across all inspected segments.

For defects d_i an d_j, the co-occurrence count is defined as:

$$C_{ij} = \left|\{p \in P : d_i \in p \wedge d_j \in p\}\right| \quad (1)$$

where P represents the set of all inspected pipe segments, D_p is the defect profile of the segment p, and $|.|$ denotes set cardinality. This count represents the number of segments in which both defects d_i and d_j are simultaneously present. The defect co-occurrence network is constructed as a weighted, undirected graph:

$$G = (V, E, W) \quad (2)$$

where $V = D$ represents nodes corresponding to unique defect types, $E \subseteq V \times V$ represents edges connecting defect pairs that co-occur in at least one segment, and $W : E \rightarrow \mathbb{R}^+$ assigns weights quantifying co-occurrence strength. Specifically, an edge $e_{ij} \in E$ exists between nodes d_i and d_j if $C_{ij} > 0$, with corresponding weight:

$$w_{ij} = C_{ij} \quad (3)$$

where C_{ij} is defined in Equation (1)

To illustrate the network construction logic, consider a simplified example involving three defects: crack (CR), fracture (FR), and root intrusion (RI), co-occurring across three inspected segments with varying frequencies. Segment analysis reveals that CR and FR appear together in two segments ($C_{CR,FR} = 2$), FR and RI co-occur in two segments ($C_{FR,RI} = 2$), while CR and RI co-occur in only one segment ($C_{CR,RI} = 1$). The resulting network contains three nodes (CR, FR, RI) connected by edges weighted 2, 2, and 1, respectively, capturing the relative strength of pairwise associations. The raw frequency weighting directly reflects the strength of association between defect pairs across the inspected network. However, to assess sensitivity to weighting scheme selection, we systematically evaluate four alternative methods that normalise co-occurrence based on different baseline expectations, including Jaccard similarity, Pointwise mutual information (PMI), lift, and normalised co-occurrence, as detailed in the sensitivity validation section.

3.2. Defect cluster identification

The identification of deterioration mechanisms from the defect co-occurrence network requires systematic partitioning of defects into clusters, groups of defects that exhibit stronger internal associations than connections to defects in other groups. We employ the Louvain algorithm (Blondel et al., 2008), a hierarchical optimisation method widely adopted in network science for its computational efficiency and ability to identify multi-scale cluster structure in large weighted networks (Que et al., 2015).

3.2.1. Louvain algorithm (LA)

A network partition assigns each defect to exactly one cluster. Let C_i denote the cluster assignment of defect i, such that $C_i \in \{1, 2, ..., K\}$ where K is the number of clusters in the partition. The Louvain operates through iterative optimisation of modularity Q, a quality metric quantifying the strength of cluster structure:

$$Q = \frac{1}{2m}\sum_{ij}\left[w_{ij} - \gamma\frac{k_i k_j}{2m}\right]\delta(c_i, c_j) \quad (4)$$

where w_{ij} represents the edge weight between defects i and j, $k_i = \sum_j w_{ij}$ is the weighted degree (total connection strength) of the defect i, $m = \frac{1}{2}\sum_{ij} w_{ij}$ is the total edge weight in the network, c_i denotes the cluster assignment of the defect i, and $\delta(c_i, c_j) = 1$ if $c_i = c_j$ and 0 otherwise. Modularity compares the fraction of edge weight within clusters to the expected fraction if edges were randomly distributed while preserving node degrees. Values approaching 1 indicate a strong cluster structure, while $Q \approx 0$ suggests random connectivity patterns (Newman & Girvan, 2004). The Louvain algorithm proceeds in two alternating phases repeated until convergence:

Phase 1: Local Optimization. Each defect is initially assigned to its own cluster. The algorithm then considers each defect i sequentially and evaluates the modularity gain, ΔQ from moving it from its current cluster to the cluster of each neighbour j:

$$\Delta Q = \left[\frac{\sum_{in} + 2k_{i,in}}{2m} - \left(\frac{\sum_{\text{tot}} + k_i}{2m}\right)^2\right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{\text{tot}}}{2m}\right)^2 - \left(\frac{k_i}{2m}\right)^2\right] \quad (5)$$

where $\sum_{in}$ is the sum of edge weights within the target cluster, $\sum_{tot}$ is the sum of edge weights incident to defects in the cluster, k_i is the weighted degree of defect i, and $k_{i,in}$ represents edge weights from i to defects in the target cluster. Defect i is moved to the cluster, yielding maximum positive ΔQ , or remaining in its current cluster if no positive gain exists. This process repeats until no individual move improves modularity.

Phase 2: Network Aggregation. Identified clusters are contracted into super nodes, creating a new network in which each node represents a cluster from Phase 1. Edge weights between super-nodes equal the sum of weights between defects in the respective clusters, while self-loops capture internal cluster connectivity. This aggregated network becomes the input for the next iteration. The algorithm terminates when modularity cannot be further improved, yielding a hierarchical decomposition where each level represents progressively coarser cluster structure. For deterioration mechanism identification, we extract the partition from the level achieving maximum modularity, as this represents the optimal balance between granularity and meaningful grouping.

3.2.2. Application to defect network

Applying the Louvain algorithm to the weighted defect co-occurrence network $G = (V, E, W)$ identifies clusters of defects that co-occur more frequently with each other than with defects outside the cluster. Each identified cluster is interpreted as a candidate deterioration mechanism, a set of defects likely driven by shared physical processes or representing sequential manifestations of the same underlying cause. For instance, a cluster with longitudinal cracks, circumferential fractures, and deformation may represent structural deterioration due to soil loading, whereas root intrusion, joint displacement, and infiltration may indicate groundwater-driven deterioration. Implementation used Python 3.9.7 with NetworkX 2.6.3 for network construction and Louvain clustering, employing a fixed random seed (random_state = 42) to ensure reproducibility. Statistical tests were conducted using SciPy 1.7.3. The Louvain algorithm offers several advantages for this application: (1) it naturally handles weighted networks, preserving infor-

mation about co-occurrence strength; (2) no prior specification of cluster number is required, allowing data-driven mechanism discovery; and (3) computational complexity of $O(nlogn)$ enables analysis of large drainage networks. The validity of identified clusters is established through multiple complementary approaches, as network-derived clusters represent statistical patterns that require verification against physical principles, domain knowledge, and predictive utility. **Section 3.3** details the validation framework, while **Section 3.4** describes the mechanism-feature association analysis used to characterise identified mechanisms.

3.2.3. Stability analysis

The Louvain algorithm implicitly assumes a resolution parameter $\gamma = 1.0$ in standard modularity optimisation. However, varying γ reveals mechanisms at different organisational scales: lower values favour fewer, larger clusters representing broad deterioration categories, while higher values yield more granular partitions that capture specific failure modes. We perform cluster detection across $\gamma \in [0.5-1.5]$ to assess whether identified mechanisms represent stable features of the network structure. For each resolution value, we apply the Louvain algorithm with a consistent random seed (42) and record the resulting partition C^{γ}. We quantify partition stability using Adjusted Mutual Information (AMI):

$$AMI = \frac{[MI(U,V) - E[MI(U,V)]]}{[max(H(U),H(V)) - E[MI(U,V)]]} \tag{6}$$

where U and V represent clustering results from different parameter settings, MI is mutual information, H denotes entropy, and E[·] represents expected value. AMI values range from 0 (no agreement) to 1 (perfect agreement), with higher values indicating greater clustering stability. The final resolution parameter is selected as the value yielding the highest modularity score across the evaluation range.

3.2.4. Defect centrality analysis

To identify structurally important defects within the co-occurrence network, three complementary centrality metrics are computed. Degree centrality $C_D(i) = \frac{k_i}{n-1}$ measures the proportion of other defects with which the defect i co-occurs, k_i is the number of direct connections, and n is the total number of defect types in the network. Betweenness centrality $C_B(i) = \sum_{s\neq i, s\neq t, t\neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$ identifies defects lying on the shortest paths between other pairs. Where σ_{st} represents the total number of shortest paths between nodes s and t, and $\sigma_{st}(i)$ is the number of those paths passing through node i. **Closeness centrality** $\boldsymbol{C_c(i) = \frac{n-1}{\sum_j d(i,j)}}$ quantifies how quickly a defect can reach all other defects in the network. $\boldsymbol{d(i,j)}$ is the shortest path distance (number of edges) between defects i and j. We emphasize that these metrics quantify topological prominence within the statistical co-occurrence network and do not establish causal or temporal roles in the initiation of deterioration.

3.3. Validation framework

The validity of identified deterioration mechanisms requires verification that they represent genuine deterioration processes rather than artefacts of the weighting scheme used to quantify co-occurrence. A robust mechanism should persist regardless of whether associations are measured by raw frequency, normalised prevalence, or statistical dependence. Different co-occurrence quantification methods may emphasize different aspects of defect relationships: raw frequency captures absolute co-occurrence strength, normalized measures account for individual defect prevalence, and mutual information-based metrics identify unexpectedly strong associations. If identified mechanisms represent genuine physical processes, the same defects should cluster together regardless of the weighting scheme used. We evaluate the five weighting methods: raw co-occurrence (**Equation 3**), Jaccard similarity $J_{ij} = \frac{C_{ij}}{|D_i \cup D_j|}$, pointwise mutual information $\mathrm{PMI}_{ij} = \log \frac{P(d_i,d_j)}{P(d_i)P(d_j)}$, lift $L_{ij} = \frac{P(d_i,d_j)}{P(d_i)P(d_j)}$, and normalized co-occurrence $N_{ij} = \frac{C_{ij}}{\min(|D_i|,|D_j|)}$, where $|D_i|$ denotes the number of segments containing a defect d_i. For each weighting scheme $k \in \{\text{raw, Jaccard, PMI, lift, normalized}\}$, we reconstruct the network $G_k = (V, E, W_k)$ with scheme-specific edge weights, apply Louvain community detection with consistent parameters ($\gamma = 1.2$), and record the resulting partition C^k. We assess cluster membership consistency across schemes using the Adjusted Rand Index:

$$\mathrm{ARI}(C^k, C^{k'}) = \frac{\sum_{ij}\binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]/\binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]/\binom{n}{2}} \tag{7}$$

where n_{ij} is the number of defects in cluster i of the partition C^k and cluster j of the partition $C^{k'}$, $a_i = \sum_j n_{ij}$ and $b_j = \sum_i n_{ij}$ are marginal sums, and n is the total number of defects. ARI ranges from −1 to 1, with 1 indicating identical clustering, 0 indicating random agreement, and negative values indicating less agreement than expected by chance. High ARI values across pairwise comparisons of the scheme demonstrate that the identified mechanisms are robust to weighting-method selection, validating that they capture fundamental patterns of defect co-occurrence.

3.4. Mechanism-Feature association analysis

After identifying and validating deterioration mechanisms, we investigated their associations with pipe infrastructure characteristics to support risk-based asset management. Multi-defect entries were parsed and disaggregated, with each defect type assigned exclusively to its corresponding cluster (1–4) based on the Louvain partition. Mechanism prevalence was calculated as the percentage of defect instances belonging to a specific cluster. For instance, all defective lining and deformation instances in the dataset are assigned to a cluster created by the network analysis.

We then examine nine pipe attributes: material type, diameter (mm), age (years), length (m), total population served, traffic intensity, district, road type, and land use. These features are classified as categorical (material, district, road type, land use) or continuous (diameter, age, length, population, traffic), requiring different statistical approaches.

For continuous features, we use the Kruskal-Wallis H test to compare distributions between pipes that exhibit each mechanism and those that do not. The test statistics are:

$$H = \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(N+1) \tag{8}$$

where N is the total sample size, n_i is the sample size for group i (mechanism present or absent), and R_i is the sum of ranks for the group i. Under the null hypothesis of identical distributions, H follows a chi-square distribution with 1 degree of freedom. For significant associations, we report epsilon squared as the effect size measure: $\epsilon^2 = \frac{H}{N-1}$, where $\epsilon^2 < 0.01$ indicates a negligible effect, $0.01 \le \epsilon^2 < 0.06$ a small effect, $0.06 \le \epsilon^2 < 0.14$ a medium effect, and $\epsilon^2 \ge 0.14$ a large effect, along with median differences for practical interpretation. For categorical variables, the Chi-square test examines independence:

$$\chi^2 = \frac{\sum_{ij}(O_{ij} - E_{ij})^2}{E_{ij}} \tag{9}$$

where O_{ij} and E_{ij} represent observed and expected frequencies, respectively, under the null hypothesis that mechanism prevalence is independent of the feature category. For significant associations, we report

Cramér's V as the effect size: $V = \sqrt{\frac{\chi^2}{N(k-1)}}$, where V < 0.10 indicates a negligible association, 0.10 ≤ V < 0.20 small, 0.20 ≤ V < 0.40 medium, and V ≥ 0.40 large. Given the K identified mechanisms and 9 features, we perform 9K statistical tests. To control family-wise error rate and avoid false discoveries, we apply the Bonferroni correction with an adjusted significance threshold $\alpha_{\text{corrected}} = \frac{0.05}{9K}$. Only associations with $p < \alpha_{\text{corrected}}$ are considered statistically significant.

4. Results

4.1. Defect co-occurrence and centrality impact

The analysed dataset comprised 697 sewer pipe segments containing 1,415 documented defect instances from Hong Kong's CCTV inspection records. The majority of segments (677, 97.1%) contained exactly two concurrent defects, 19 segments (2.7%) exhibited three defects, and one segment (0.1%) recorded four simultaneous defects. The mean number of defects per segment was 2.03 (SD = 0.18), indicating consistent co-occurrence patterns with low variance across the network.

Before identifying distinct defect clusters using Louvain clustering, we first examined pairwise co-occurrence patterns and network centrality metrics to reveal the fundamental structural relationships among defect types. These centrality metrics quantify topological prominence within the statistical co-occurrence network, specifically, breadth of connectivity across diverse deterioration contexts, and do not establish causal or temporal roles in deterioration initiation. The co-occurrence matrix (Fig. 3) reveals distinct clustering patterns among the 15 defect types, with three dominant defect hubs emerging: Encrustation-Fracture (70 co-occurrences), Defective Lining-Deformation (125 co-occurrences), and Broken-Infiltration (57 co-occurrences). Notably, the Defective Lining-Deformation pair exhibited the strongest co-occurrence relationship, suggesting that these defects represent a unified deterioration mechanism. Beyond these pairwise relationships, Encrustation demonstrated the broadest connectivity pattern, co-occurring substantially with Broken (54), Fracture (70), and Infiltration (52), positioning it as the most broadly connected node in the co-occurrence network. This pattern is confirmed by its highest degree centrality (0.857) and betweenness centrality (0.178) as shown in Table 1, indicating its role as a topological hub within the statistical co-occurrence structure; whether this reflects causal primacy, diagnostic sensitivity, or prevalence effects requires longitudinal validation (see

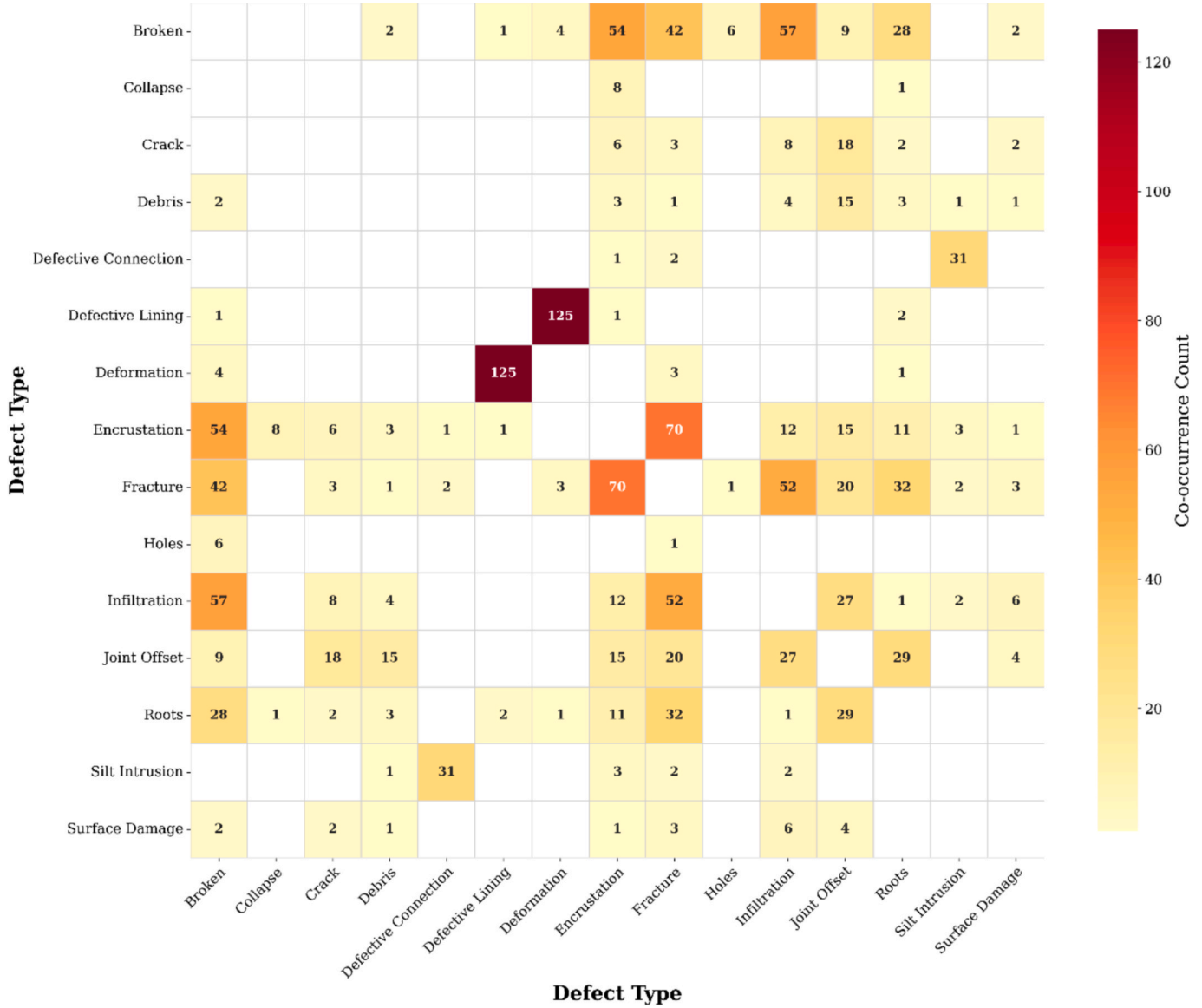

Fig. 3. Defect Co-Occurrence Matrix.

Table 1
Centrality Metrics Summary.

Defect	Total Co-occurrences	Degree Centrality	Betweenness Centrality	Closeness Centrality
Broken	205.0	0.714	0.095	0.778
Collapse	9.0	0.143	0.000	0.519
Crack	39.0	0.429	0.002	0.636
Debris	30.0	0.571	0.013	0.700
Defective Connection	34.0	0.214	0.000	0.560
Defective Lining	129.0	0.286	0.003	0.583
Deformation	133.0	0.286	0.003	0.583
Encrustation	185.0	0.857	0.178	0.875
Fracture	231.0	0.857	0.178	0.875
Holes	7.0	0.143	0.000	0.519
Infiltration	169.0	0.643	0.020	0.737
Joint Offset	137.0	0.571	0.005	0.700
Roots	110.0	0.714	0.097	0.778
Silt Intrusion	39.0	0.357	0.007	0.609
Surface Damage	19.0	0.500	0.004	0.667

Section 5.2d).

Similarly, Fracture exhibited comparable topological importance (degree centrality = 0.857) with strategic connections to structural defects, including Broken (42), Infiltration (52), and Roots (32). Conversely, Collapse (9 total co-occurrences) and Holes (7 co-occurrences) exhibit minimal network integration (betweenness centrality = 0.000), indicating isolated occurrence patterns potentially linked to specific failure mechanisms or localised conditions rather than systemic deterioration processes. Both Holes and Collapse are considered ultimate failure points; it is expected that only a few defects are associated with them.

4.2. Defect clustering

While the co-occurrence matrix and centrality analysis reveal pairwise relationships and individual defect importance, identifying meaningful defect clusters requires efficient clustering analysis across multiple resolution parameters to capture both fine-grained and broader deterioration patterns.

4.2.1. Parameter optimisation and defect clusters stability

The sensitivity analysis reveals critical insights into parameter selection for robust defect clustering. Fig. 4**(a)** demonstrates that the network structure undergoes a substantial transformation as thresholds increase, edge count declines from 51 to 17, while network density decreases sharply from 0.49 to 0.16. Despite this sparsification, network modularity stabilizes around 0.35-0.40 beyond threshold 4, as illustrated in Fig. 4**(a)**. Resolution parameter analysis (Table 2) further shows that modularity reaches peak values ($Q \approx 0.337$) at $\gamma = 1.0$-1.2, confirming a stable,defined cluster structure at the selected resolution parameter.

For resolution parameter optimisation, Fig. 4**(b)** shows that resolution 1.0 produces three balanced clusters with peak modularity (0.337)

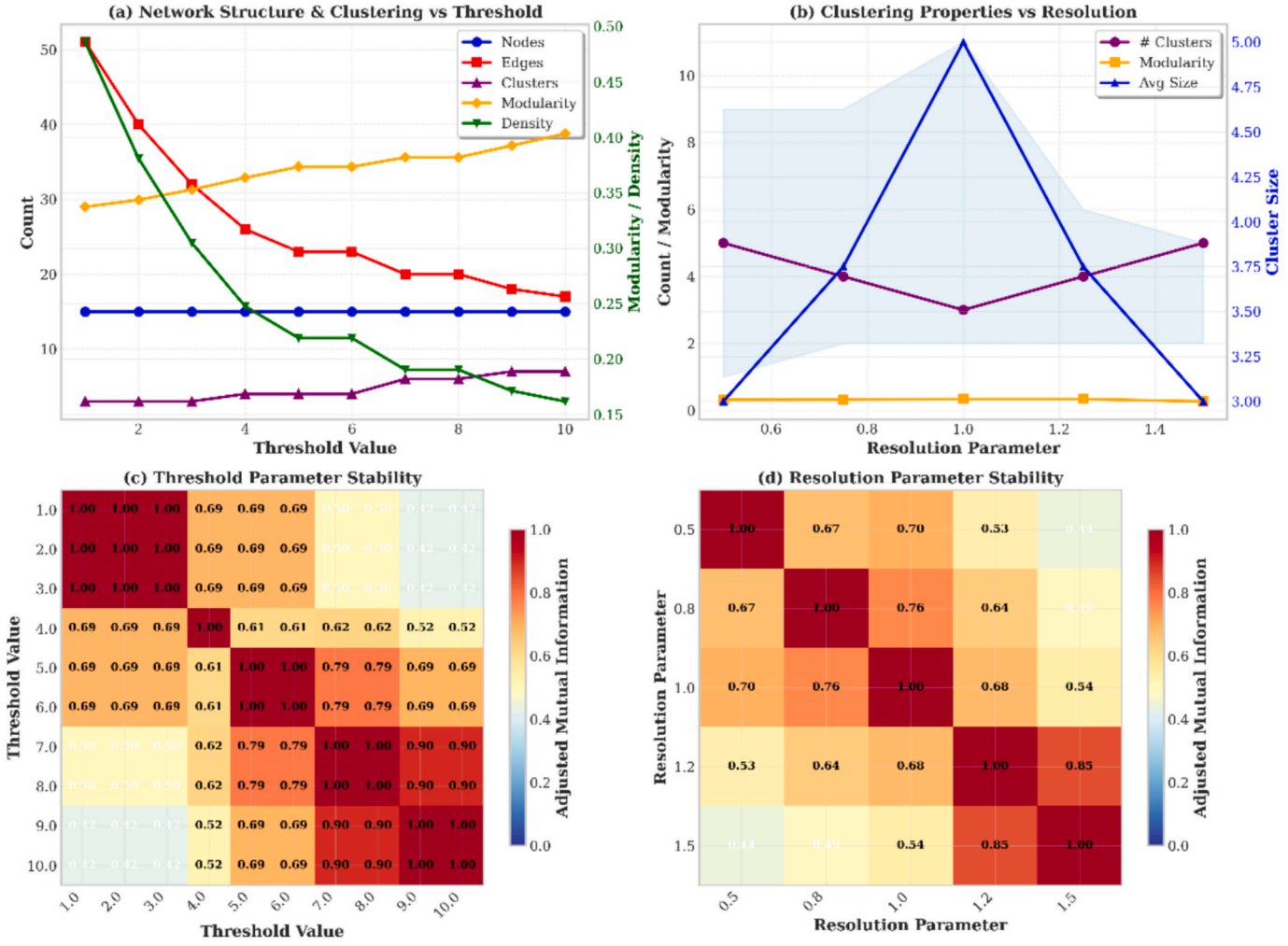

Fig. 4. Threshold and Resolution Sensitivity.

Table 2
Network Metrics Across Resolution Parameters.

Resolution (γ)	Num. Clusters	Modularity	Max Size	Min Size	Avg Size	Std Size
0.5	4	0.3239	10	1	3.75	3.63
0.6	4	0.3239	10	1	3.75	3.63
0.7	4	0.3239	10	1	3.75	3.63
0.8	3	0.3373	11	2	5.00	4.24
0.9	4	0.3314	7	2	3.75	2.05
1.0	3	0.3373	11	2	5.00	4.24
1.1	4	0.3372	6	2	3.75	1.79
1.2	4	0.3372	6	2	3.75	1.79
1.3	4	0.3372	6	2	3.75	1.79
1.4	4	0.3372	6	2	3.75	1.79
1.5	5	0.2677	5	2	3.00	1.10

and a mean cluster size of 5.0. However, to capture more granular defect relationships, we select resolution 1.2, which identifies four clusters while maintaining high modularity (0.337) and more uniform cluster sizes (mean = 3.75, std = 1.79). The stability analysis validates this choice: Fig. 4**(c-d)** demonstrates strong parameter robustness, with threshold values 4–8 showing high agreement (AMI > 0.79) and resolution parameters 0.8–1.2 exhibiting consistent clustering (AMI > 0.64). Therefore, we adopt threshold = 4 and resolution = 1.2 as optimal parameters, balancing network connectivity with granular cluster detection. Subsequently, Section 4.2.2 will evaluate different edge weighting schemes to further validate and refine these clustering results.

Based on the network visualization (Fig. 5), four distinct clusters were identified. The following interpretation provides insights into the structural patterns of the defect co-occurrence network.

4.2.1.1. Cluster 1 (n = 2, density = 1.000): defective lining, deformation. This cluster represents severe structural compromises with complete internal connectivity (density = 1.000), indicating strong co-occurrence between deformation and lining failures. These defects suggest a failure-cascade mechanism in which external loading causes deformation that damages protective linings, or, conversely, deteriorated linings expose pipes to corrosive environments, weakening structural integrity. Both defects require immediate structural rehabilitation, as they threaten the overall integrity of the pipe.

4.2.1.2. Cluster 2 (n = 6, density = 0.600): broken, collapse, encrustation, fracture, holes, infiltration. This largest cluster captures progressive physical deterioration pathways from minor fractures to complete collapse. The moderate density (0.600) reflects complex interdependencies in which encrustation accelerates structural failures by increasing hydraulic stress, while broken sections and holes facilitate infiltration, leading to soil erosion and loss of external support. This cluster demonstrates critical hydraulic-structural coupling requiring condition-based maintenance strategies.

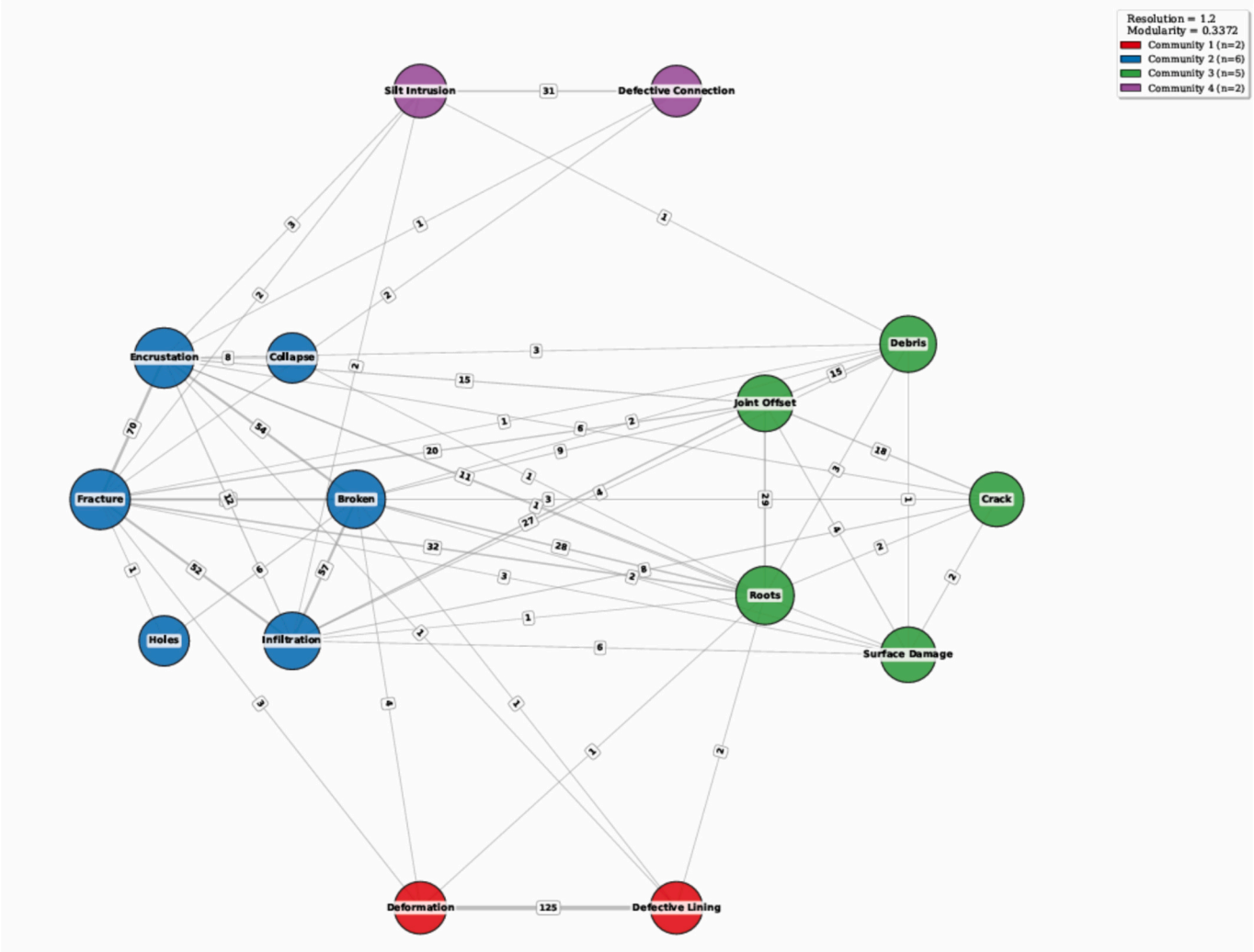

Fig. 5. Optimal Co-occurrence Network at resolution 1.2.

4.2.1.3. Cluster 3 (n = 5, density = 0.800): crack, debris, joint offset, roots, surface damage. With high internal connectivity (density = 0.800), this cluster represents joint-centric failures and environmental intrusions. Joint offsets create flow disruptions, causing debris accumulation and stress concentrations that initiate cracks, while root intrusion through existing openings progressively damages surfaces and displaces joints.

4.2.1.4. Cluster 4 (n = 2, density = 1.000): defects: defective connection, silt intrusion. This isolated cluster with perfect internal connectivity represents infrastructure interface failures at pipe junctions, service laterals, and manhole connections. Defective sealing at connection points enables silt intrusion, providing a statistically associated indicator of groundwater infiltration. These defects require targeted sealing interventions at connection points rather than full pipe replacement.

4.2.2. Clustering validation using alternative weighting method

The clusters identified above reveal distinct failure mechanisms ranging from critical structural compromises to joint-specific deterioration. To assess whether the identified structure represents genuine co-occurrence patterns rather than frequency bias, alternative weighting schemes were compared. A sensitivity analysis using alternative weighting schemes was conducted to assess clustering stability and robustness.

Five edge weighting methods were evaluated: Raw (frequency-based), Jaccard (similarity coefficient), PMI (pointwise mutual information), Lift (association strength), and Normalised (proportional weighting). Table 3 summarises the clustering outcomes, while Fig. 6(a) and (b) illustrate the relationship between the resolution parameter, the number of clusters, and the modularity scores across all weighting schemes. The majority of weighting schemes (Raw, Jaccard, Lift) consistently identified four clusters at resolution 1.2, demonstrating strong structural stability. Both PMI and Normalised methods produced five clusters, suggesting that these approaches detect finer-grained cluster structures by emphasising different aspects of defect relationships.

As shown in Fig. 6**(b)**, PMI achieved the highest modularity (Q = 0.7292), indicating the strongest cluster separation, followed by Jaccard (Q = 0.5571) and Lift (Q = 0.5299). The Raw frequency-based approach yielded the lowest modularity (Q = 0.3372), confirming that simple co-occurrence counts may underestimate true cluster structure. Notably, PMI's modularity remains consistently high (>0.70) across all resolution parameters, demonstrating superior stability.

In Fig. 7, the ARI analysis revealed strong methodological consistency in defect clustering, with perfect agreement observed within frequency-based methods (Raw-Jaccard, ARI = 1.000) and probabilistic approaches (Lift-Normalised, ARI = 1.000). Cross-method comparisons demonstrate moderate-to-good agreement (ARI = 0.438–0.748), indicating that different weighting schemes capture complementary aspects of defect co-occurrence patterns while maintaining overall structural similarity. Notably, PMI exhibits balanced performance, with moderate agreement. These findings validate the robustness of identified defect clusters across diverse weighting viewpoints.

Table 3
Weighting methods.

Scheme	Clusters	Modularity	Avg Size	Std Size
Raw	4	0.3372	3.75	1.79
Jaccard	4	0.5571	3.75	1.79
PMI	5	0.7292	3.00	1.10
Lift	4	0.5299	3.75	1.79
Normalized	5	0.4051	3.00	2.00

4.3. Cluster feature association analysis

Having established the robustness of the four-cluster defect structure through multi-method validation (ARI analysis showing moderate-to-good cross-method agreement, 0.438–0.748), the analysis now shifts from understanding which defects co-occur to investigating why these patterns emerge. This section examines the associations between the identified defect clusters and pipe-specific characteristics (material, age, diameter, length) and environmental conditions (land use, traffic loading). The dataset integrated inspection records with comprehensive pipe-level physical, operational, and environmental features spanning infrastructure characteristics (diameter, age, materials: Vitrified Clay/Concrete), spatial context (districts, land use types), and operational demands (population density, traffic volumes) (Table 4).

FromTables 5 and 6, statistical testing revealed that deterioration mechanisms exhibit distinct infrastructure profiles. Geographic factors demonstrated the strongest discriminatory power: District showed medium-to-strong effect sizes across all mechanisms, with Cluster 3 exhibiting exceptional spatial clustering. Land-use associations were consistently significant across mechanisms, whereas road-type effects varied. Material composition showed mechanism-specific patterns, with Cluster 4 exhibiting the strongest material dependence and Cluster 3 showing a weaker association. Among continuous features, age showed strong differentiation: Cluster 3 defects occurred in significantly newer infrastructure (10 years newer), while Cluster 1 concentrated in older pipes (6 years older). Pipe diameter significantly distinguished mechanisms: Cluster 4 defects occurred in large-diameter collectors (150 mm larger), Cluster 3 in mid-size trunk sewers (75 mm larger), while Cluster 1 showed no diameter preference. Segment length associations revealed operational patterns: Cluster 4 defects occurred in exceptionally long inspection segments (16 m longer), suggesting connection points in major interceptors, while Cluster 1 concentrated in shorter segments (5.5 m shorter). Population density significantly elevated Cluster 3 prevalence while reducing Cluster 1 occurrence. Traffic intensity showed marginal effects, with only Cluster 3 achieving significance, suggesting traffic loading plays a secondary role in defect mechanisms.

5. Discussion

5.1. Interpretation of defect clusters: from co-occurrence patterns to deterioration mechanisms

The identification of four robust defect clusters through multi-method validation reveals distinct co-occurrence patterns in sewer infrastructure, each characterised by unique statistical associations with pipe and environmental features. However, translating these empirical patterns into mechanistic interpretations requires careful consideration of analytical limitations and alternative explanations. While we propose reasonable deterioration pathways based on engineering principles and domain knowledge, readers should recognise that the cross-sectional nature of inspection data and network-based methodology impose important constraints on causal inference.

5.1.1. Cluster 1: Lining-deformation co-occurrence pattern (18.4% prevalence)

Cluster 1 exhibits the strongest pairwise association in the entire dataset: defective lining and deformation co-occur 125 times with perfect within-cluster connectivity (density = 1.000). This exceptional co-occurrence frequency substantially exceeds what would be expected by chance, suggesting a strong relationship between these defect types. The observed pattern represents coordinated failure of the lining system, with protective barriers and structural integrity deteriorating in tandem. The cluster, per the feature association, is characterised by older vitrified clay pipes, which indicate lining wear and potential environmental impacts from prolonged use.

Three plausible deterioration pathways could explain this co-

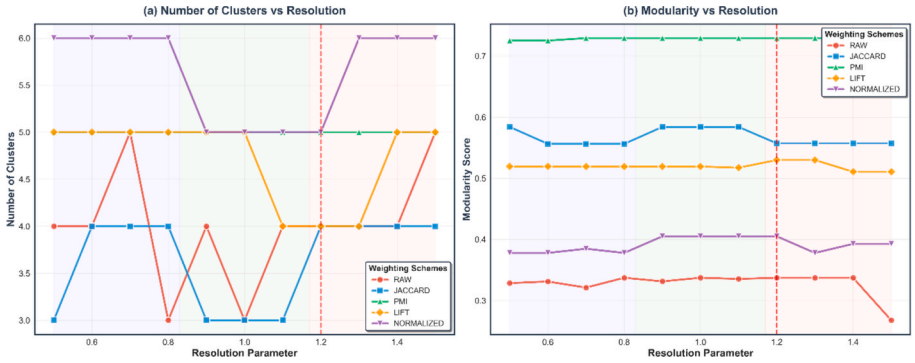

Fig. 6. Multiple Weighting Schemes for Network Edges.

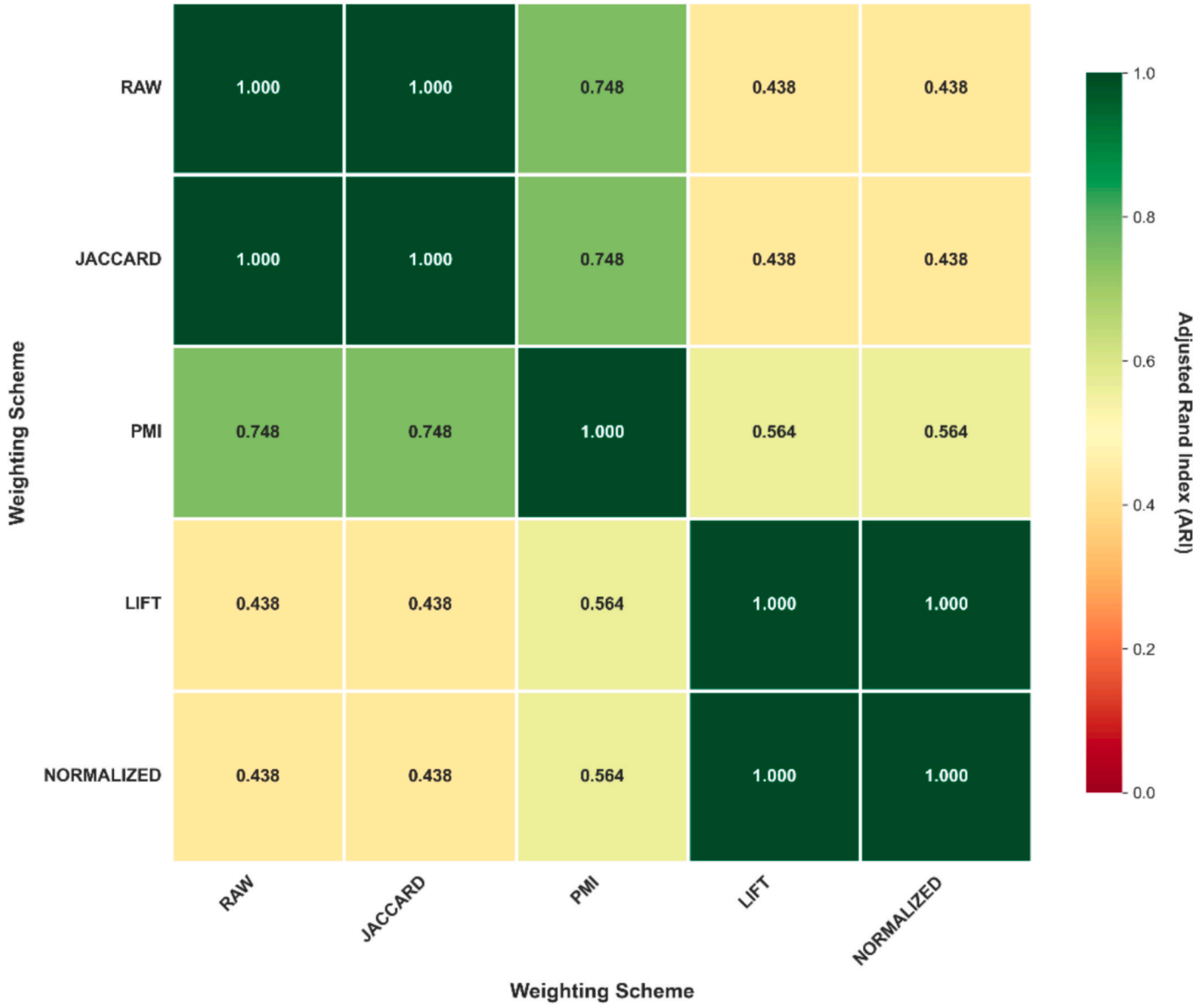

Fig. 7. Comparative weighting Agreement by ARI.

occurrence: (1) External loading causes structural deformation that mechanically damages protective linings through tensile or shear stress (Falter, 1996). The transfer of soil loading to liners is controlled by the complex interaction between the host pipe and the surrounding soil. As the host pipe deteriorates or voids develop in the soil matrix, the pipe–soil system deforms. This structural adjustment directly influences how the liner engages with the host pipe, shaping the load distribution and the liner's performance within the composite system (Spasojevic et al., 2007); (2) deteriorated linings expose pipe walls to aggressive chemical environments (sulfuric acid from microbial activity, industrial

Table 4
Dataset Feature Summary and Descriptive Statistics.

Feature	Data Type	Unique Values	Sample Values	Category	Description
Diameter	Integer	24	225, 300, 375 mm	Physical	Pipe nominal diameter (range: 150–1800 mm)
Age	Integer	67	33, 41, 28 years	Physical	Years since installation (range: 5–72 years)
Length	Float	632	28.58, 13.53, 11.23 *m*	Physical	Inspected segment length (range: 2.1–156.4 m)
Material	Categorical	2	Vitrified Clay, Concrete	Physical	Pipe construction material
District	Categorical	10	Tai Po, Yuen Long, Sai Kung	Geographic	Administrative district location
Land Use	Categorical	11	Village Type Development, Residential	Contextual	Surrounding land use classification
Road Type	Categorical	7	PD, TR, UR	Contextual	Road classification (PD=Primary Distributor, TR = Trunk Road, UR = Urban Road)
Total Population	Float	106	13412, 15906, 17,227	Operational	Service area population density
AADT	Float	294	29646.67, 18523.45	Operational	Annual Average Daily Traffic (vehicles/day)

Table 5
Continuous Features – Kruskal-Wallis H Test Results.

Feature	Mechanism	H	p-value	ε^2	Effect Size	Δ Median
Diameter	1	2.039	0.1533	0.0014	Small	0.000
	2	40.186	<0.0001	0.0284	Small	0.000
	3	22.450	<0.0001	0.0159	Small	75.000
	4	30.786	<0.0001	0.0218	Small	150.000
Age	1	14.261	0.0002	0.0101	Small	6.000
	2	8.232	0.0041	0.0058	Small	2.000
	3	51.626	<0.0001	0.0365	Small	−10.000
	4	0.000	0.9899	0.0000	Small	−4.000
Length	1	58.242	<0.0001	0.0412	Small	−5.465
	2	3.489	0.0618	0.0025	Small	0.930
	3	18.237	<0.0001	0.0129	Small	3.800
	4	36.461	<0.0001	0.0258	Small	16.135
Total Population	1	17.747	<0.0001	0.0127	Small	−831.000
	2	0.357	0.5504	0.0003	Small	−473.500
	3	40.201	<0.0001	0.0288	Small	2152.000
	4	2.364	0.1242	0.0017	Small	−749.000
AADT	1	0.000	0.9945	0.0000	Small	−144.450
	2	2.671	0.1022	0.0019	Small	108.055
	3	5.757	0.0164	0.0041	Small	1116.110
	4	2.930	0.0870	0.0021	Small	−2286.110

Table 6
Categorical Features – Chi-Square Test Results.

Feature	Mechanism	χ^2	p-value	V	Effect Size	Sig.	df
Material	1	8.562	0.0034	0.0778	Small	**	1
	2	18.269	<0.0001	0.1136	Small	***	1
	3	6.409	0.0114	0.0673	Small	*	1
	4	33.755	<0.0001	0.1545	Medium	***	1
District	1	97.978	<0.0001	0.2631	Medium	***	9
	2	102.345	<0.0001	0.2689	Medium	***	9
	3	290.415	<0.0001	0.4530	Strong	***	9
	4	80.158	<0.0001	0.2380	Medium	***	9
Road Type	1	40.203	<0.0001	0.1686	Medium	***	6
	2	20.338	0.0024	0.1199	Small	**	6
	3	69.776	<0.0001	0.2221	Medium	***	6
	4	13.816	0.0318	0.0988	Small	*	6
Land Use	1	44.860	<0.0001	0.1781	Medium	***	10
	2	30.373	0.0007	0.1465	Small	***	10
	3	51.549	<0.0001	0.1909	Medium	***	10
	4	67.749	<0.0001	0.2188	Medium	***	10

effluents), accelerating corrosion that reduces structural capacity and enables deformation; or (3) a common external cause, such as severe ground movement, heavy traffic loading, or loss of pipe bedding, simultaneously produces both defect types. (Law & Moore, 2007) determined from finite element analysis that both the host pipe and the liner deform under the full overburden pressure if the host pipe fractures after liner installation. Achieving a safe liner design, therefore, requires ensuring that local bending associated with host pipe fracture, as well as deformation under full soil loads, remains within the tensile capacity of the polymer material. The cross-sectional CCTV data cannot distinguish among these pathways, as all three predict the observed co-occurrence pattern. Longitudinal inspection data tracking pipes from the initial appearance of a defect through subsequent deterioration would be required to establish temporal precedence and validate specific causal sequences.

Practically, the perfect co-occurrence pattern (density = 1.000) indicates that inspection programs identifying either defective lining or deformation should flag affected pipes for comprehensive assessment of

both defect types. The strong association suggests that rehabilitation addressing only one defect type may prove inadequate; pipes with advanced deformation may experience rapid lining deterioration even after lining repair, while lining replacement on deformed pipes may fail prematurely due to persistent mechanical stress. Structural rehabilitation methods that simultaneously restore geometric integrity and provide chemical protection (e.g., cured-in-place pipe with corrosion-resistant resins, structural slip-lining) appear more appropriate than isolated interventions.

5.1.2. Cluster 2: Structural-hydraulic co-occurrence pattern (54.5% prevalence)

This cluster encompasses the most prevalent co-occurrence pattern, characterised by interconnected structural defects (broken, collapsed, fractured, and holes) and operational impairments (encrustation and infiltration) with moderate internal connectivity (density = 0.600). This Structural-Hydraulic Co-occurrence Pattern reflects multiple deterioration pathways likely contributing to the observed co-occurrences rather than a single linear mechanism. The network centrality analysis (Section 4.1) identified encrustation as having the highest degree centrality (0.857) and betweenness centrality (0.178) within this cluster, indicating it frequently co-occurs with multiple structural defect types. However, the high centrality of encrustation merits careful interpretation. Network centrality measures statistical association patterns; encrustation's central position indicates it frequently appears in pipes with diverse structural defects but does not directly establish temporal precedence or causal initiation. This cluster could also have three interpretations: (1) encrustation develops early in deterioration sequences and subsequently contributes to structural failures through increased hydraulic stress (Yusuf et al., 2022); (2) Minor structural defects (small fractures, surface roughness) create flow disturbances and retention points that promote encrustation accumulation, making encrustation a consequence rather than a cause (Goode et al., 2025; Yang et al., 2026); or (3) common underlying conditions (low flow velocities, aggressive water chemistry, sediment loading) produce both encrustation and structural defects independently, with no direct causal relationship between them. This cluster is characterised by balanced concrete and vitrified clay pipes and is subject to heavy traffic and longevity requirements. This heavy traffic explains the prevalence of crucial structural defects in this cluster.

Engineering principles support multiple mechanisms: encrustation reduces effective pipe diameter, increasing flow velocity and wall shear stress on the remaining cross-section, which could accelerate structural degradation. Additionally, encrustation layers can trap moisture and aggressive chemicals against pipe walls, potentially accelerating corrosion beneath deposits. Distinguishing these alternatives requires longitudinal data tracking defect appearance sequences. If encrustation systematically precedes structural defects in affected pipes, earlier causation gains support, and if structural defects appear first, the consequence interpretation is more plausible. For practice, the high prevalence (54.5%) indicates this pattern represents the dominant challenge requiring systematic condition-based management. Despite interpretive uncertainties about causal sequences, the strong statistical association between encrustation and structural defects provides actionable guidance: pipes exhibiting significant encrustation warrant elevated monitoring priority even in the absence of visible structural defects, as the co-occurrence pattern indicates elevated risk. Whether encrustation serves as a catalyst, an indicator, or a co-effect, its presence reliably signals pipes requiring attention.

5.1.3. Cluster 3: Root-joint co-occurrence pattern (22.3% prevalence)

Cluster 3 exhibits high internal connectivity (density = 0.800) among joint offset, roots, debris, crack, and surface damage. This Root-Joint Co-occurrence Pattern pattern emphasises the spatial co-location around pipe joints and environmental intrusion. The co-occurrence patterns suggest, but do not definitively establish, that joint vulnerability serves as an organising principle: joint offsets create flow disruptions promoting debris accumulation, while joint gaps provide entry points for root intrusion. According to (Randrup et al., 2001), roots are reported to cause > 50% of all sewer blockages, and the costs associated with root removal from sewers are substantial. Kielce University of Technology collected data on root intrusions into sewers, along with the defects that facilitate such intrusions. They determined the category of root intrusion consequences for sewers with roots already growing inside, and a method for determining the sewage flow blockage risks related to root intrusion (Kuliczkowska & Parka, 2017). However, alternative sequences remain probable: root intrusion through initially minor joint imperfections could cause progressive displacement leading to observable joint offset, or differential settlement could simultaneously produce both joint misalignment and soil cracking that enables root penetration (Vipulanandan & Liu, 2005).

Practically, the extreme spatial clustering enables geographically targeted interventions with high efficiency. Districts exhibiting elevated prevalence should receive a comprehensive investigation to identify systematic installation deficiencies or environmental risk factors (e.g., soil shrink-swell potential, tree canopy density, vibration exposure). The high within-cluster connectivity (density = 0.800) suggests that addressing individual defect types without comprehensive joint rehabilitation provides limited benefit, as the interconnected nature of defects indicates persistent joint vulnerability. Preventive measures targeting root intrusion (chemical root barriers, species-specific tree management) may be more cost-effective than reactive mechanical root removal if joint gaps serve as entry points. Also, this cluster is prevalent in comparably newer pipes and more populated residential areas. They represent early-stage defects and require proactive solutions before they develop into failures.

5.1.4. Cluster 4: Connection-sediment co-occurrence pattern (4.9% prevalence)

Cluster 4 exhibits perfect co-occurrence (density = 1.000) between defective connections and silt intrusion, with 31 documented instances. This connection-point sediment intrusion pattern reflects the spatial specificity of infrastructure interfaces (service laterals, manhole connections, junction structures) rather than continuous pipe segments. The perfect co-occurrence likely reflects the physical mechanism by which failed seals at connection points permit groundwater infiltration carrying suspended fine sediments. Unlike the coarser debris accumulation in Cluster 3 (associated with flow disruptions), the silt composition indicates external water ingress rather than internal system materials. In practice, despite a low prevalence (4.9%), connection points warrant disproportionate attention due to their strategic importance; service lateral failures affect individual properties directly; junction failures can compromise network connectivity; and connection-related infiltration disproportionately contributes to inflow and infiltration (I&I) problems. The perfect co-occurrence between defective connections and silt intrusion provides a low-cost screening indicator: routine manhole inspections documenting sediment accumulation patterns can identify connection defects without the expense of CCTV deployment. The strong material and dimensional associations enable risk-based prioritisation focusing on large-diameter collectors with connection-intensive configurations in vulnerable material types.

5.2. Methodological considerations and limitations

(a) Temporal Dynamics and Cross-Sectional Constraints:

The most significant limitation of this analysis is treating defect relationships as static, despite the dataset spanning 14 years (2007–2021). Cross-sectional CCTV data capture defect co-occurrence at single time points, but do not observe temporal sequencing within deterioration pathways. Consequently, our mechanistic interpretations, while grounded in engineering principles and consistent with observed co-occurrence patterns, remain fundamentally inferential rather than

empirically validated causal models. The network analysis identifies which defects frequently co-occur but cannot establish which defects appear first in deterioration sequences or whether co-occurring defects share causal relationships or merely respond to common underlying drivers. For example, Cluster 2′s Structural-Hydraulic Co-occurrence Pattern pattern could follow multiple sequences: fracture → encrustation → broken → infiltration → collapse represents one plausible pathway consistent with observed co-occurrences, but alternative sequences (infiltration → soil erosion → external load increase → structural failure) could produce identical co-occurrence patterns. Longitudinal analysis tracking individual pipe segments through multiple inspection cycles could validate specific failure sequences by establishing temporal precedence.

Future research should incorporate time-to-event analysis (survival models, Cox proportional hazards) to estimate deterioration progression rates and identify factors accelerating or retarding transitions between defect states. The 14-year observation window provides sufficient temporal depth for such analysis if inspection records are linked longitudinally at the pipe segment level rather than analysed cross-sectionally. Additionally, dynamic network models that explicitly represent the temporal evolution of defect co-occurrences could reveal whether deterioration pathways have changed over time due to shifting demographics, climate change, or evolving pipe material compositions in the active inventory.

(b) Edge Weighting Methodology and Sensitivity Analysis:

The primary network analysis employed raw co-occurrence frequency as edge weights, potentially introducing bias toward highly prevalent defects. For example, encrustation's high degree of centrality (0.857) and betweenness centrality (0.178) partially reflect its overall high prevalence (185 total co-occurrences) rather than uniquely strong mechanistic relationships. A defect that co-occurs randomly with all other defects, proportional to their prevalence, would still achieve high centrality simply through frequency effects. Section 4.2.2 evaluated four alternative weighting schemes that normalise for base prevalence: Jaccard, PMI, Lift, and Normalised. The key finding is that the four-cluster structure remained largely stable across weighting methods. However, modularity scores varied substantially, with PMI achieving a Q value of 0.7292. Future analyses should consider PMI or similar normalised metrics as primary weighting schemes, with raw frequency providing supplementary context about absolute co-occurrence magnitudes.

(c) Cluster Validation and External Corroboration:

The use of Adjusted Mutual Information (AMI) effectively demonstrates clustering stability across parameter ranges, threshold values 4–8 maintain high agreement (AMI $>$ 0.79), and resolution parameters 0.8–1.2 exhibit consistent structure (AMI $>$ 0.64), confirming that identified clusters are robust. However, AMI represents internal validation assessing consistency across algorithmic variants, not external validation confirming correspondence to real-world failure processes. Several approaches could strengthen external validation: (1) Failure case comparison: Retrospective analysis of documented catastrophic failures to determine whether pre-failure inspection records exhibit defect combinations matching identified clusters. (2) Expert elicitation: Structured interviews with experienced utility engineers and CCTV analysts to assess whether proposed deterioration pathways align with field observations and professional judgment accumulated over decades of infrastructure management. (3) Hydraulic-structural modelling: Computational simulation of deterioration processes to determine whether proposed mechanisms are physically plausible under realistic loading and environmental conditions.

(d) Centrality Interpretation: Association vs. Causation:

The characterisation of high-centrality defects requires careful qualification. Centrality measures quantify the topological importance of nodes within co-occurrence networks. Encrustation's high betweenness centrality (0.178) indicates it frequently appears on shortest paths between other defect types in the network structure, but topological centrality does not equate to causal or temporal primacy in deterioration sequences. Three distinct interpretations of centrality are possible: (1) Causal centrality: High-centrality defects occur early in deterioration sequences and directly contribute to subsequent defect development. (2) Diagnostic centrality: High-centrality defects serve as sensitive indicators of deterioration activity, appearing in pipes with diverse defect combinations because they respond to multiple underlying stressors. (3) Confounded centrality: High-centrality defects appear frequently in deteriorating pipes due to high base prevalence combined with nonspecific association with many deterioration drivers, without unique causal or diagnostic importance. Only longitudinal data establishing temporal sequences can distinguish these interpretations. For practical application, high-centrality defects warrant a higher monitoring priority because their statistical association with multiple other defect types indicates a higher risk, whether they serve as causes, consequences, or co-effects. The actionable insight that pipes with encrustation should receive increased attention remains valid across all three interpretations, though the optimal timing of the intervention differs (proactive if causal, reactive if consequential).

5.3. Integrated asset management implications

The validated cluster structure enables meaningful improvements to current asset management practices, which typically treat all infrastructure as homogeneous and apply uniform inspection and rehabilitation protocols. The four clusters exhibit distinct prevalence rates (54.5%, 22.3%, 18.4%, 4.9%), feature associations, and spatial distributions, indicating that failure patterns are fundamentally heterogeneous across the pipe inventory. The analysis has the following practical implications:

1. The Lining-Deformation co-occurrence pattern needs an age-based prioritisation ($>$40 years) combined with material targeting (Vitrified Clay) and geographic focus (urban core districts) to identify high-risk populations. Also, comprehensive solutions that address both structural integrity and environmental protection are needed. Trenchless structural methods (CIPP with structural-grade resin, close-fit slip-lining) restore both geometric stability and material protection simultaneously. Essentially, the categorical associations (material, district) with age as a continuous predictor suggest classification tree approaches (random forests, gradient boosted trees) will outperform linear models in their predictive modelling

2. Structural-Hydraulic co-occurrence Pattern, which is prevalent in ageing (40.9 years) pipes, supports condition-based monitoring where initial inspection establishes a baseline with adaptive re-inspection intervals based on severity. Pipes exhibiting significant encrustation warrant accelerated reinspection despite the absence of current structural defects, given strong co-occurrence patterns. Also, hydraulic modelling outputs (flow velocity, sediment transport capacity) will supplement inspection data to predict encrustation susceptibility. For rehabilitation, early-stage defects (minor fractures, moderate encrustation) will respond to lower-cost interventions (spot repairs, hydraulic cleaning), reserving expensive techniques (excavation, full replacement) for advanced deterioration. Modelling this cluster with multiple significant continuous predictors (age, length, population density) and categorical predictors (road type, land use) suggests that gradient-boosting models with interaction terms (age $\times$ material) will improve predictive accuracy beyond univariate approaches

3. Root-Joint co-occurrence pattern is predominant in relatively new infrastructure (34.4 years), serving the highest population densities (17,131) with elevated traffic volumes (AADT: 19,543). They may require spatial clustering (district V $=$ 0.453) to enable geographic targeting, concentrating inspection resources in affected districts (Southern: 24.1%, Yuen Long: 22.5%, Tuen Mun: 14.6%) while reducing frequency elsewhere. Combined protocols (root removal, joint grouting, and chemical root barriers) can address the interconnected defect assembly. For this cluster, spatial models (geographically weighted regression, spatial autoregressive models) will substantially outperform

non-spatial approaches by capturing localized installation quality or environmental factors not represented in available features

4. Connection-sediment co-occurrence pattern is also common in the largest diameter pipes (463.0 mm), and overwhelming concrete dominance (87.0%). Due to low prevalence, routine manhole inspections documenting sediment accumulation provide low-cost screening, with targeted CCTV lateral launches deployed where silt intrusion is observed. Material-specific prioritisation (concrete pipes with vulnerable connection technologies) and land-use targeting (industrial zones, special-use areas) improve efficiency

5.4. Future research directions

Several research extensions would address identified limitations and strengthen practical application:

1. Longitudinal deterioration tracking: Link inspection records across multiple cycles at the individual pipe segment level to establish temporal sequences, validate proposed failure pathways, and quantify progression rates between deterioration stages. This enables evidence-based estimation of optimal re-inspection intervals and intervention timing.

2. Integration of unmeasured variables: Incorporate construction quality metrics (contractor identity, bedding material specifications, installation depth), maintenance history (cleaning frequency, prior repairs), and hydraulic loading data (flow velocity, velocity: Capacity ratio, hydrogen sulphide concentration) to improve predictive model accuracy beyond the modest effect sizes achieved with available features.

3. Forensic failure investigation: Conduct structured retrospective analysis of documented failures (emergency repairs, collapses, service disruptions) to determine whether pre-failure inspection records exhibit cluster-specific defect combinations, providing external validation of mechanistic relevance

4. Rehabilitation effectiveness evaluation: Track long-term performance of different rehabilitation techniques applied to cluster-specific defect patterns to enable evidence-based treatment selection. Current rehabilitation decisions largely rely on vendor claims and engineering judgment; comparative effectiveness research stratified by failure pattern would enable optimisation of capital investments

5. Hydraulic-structural modelling: Develop computational models simulating proposed deterioration mechanisms (encrustation effects on wall shear stress, root intrusion forces on joint displacement, infiltration-induced soil erosion) to establish the physical plausibility of hypothesised failure pathways and identify critical thresholds at which deterioration accelerates

6. Generalisation assessment: Apply an identical analytical framework to inspection data from other utilities, climate zones, and demographic contexts to determine whether identified clusters represent universal failure patterns or context-specific phenomena. A multi-utility comparison would enable the identification of transferable vs system-specific insights.

6. Conclusion

This study addressed critical limitations in drainage infrastructure condition assessment by developing a network science framework that identifies systematic deterioration mechanisms from defect co-occurrence patterns. Analysis of 14 years of CCTV inspection data revealed four robust defect clusters representing distinct failure pathways: Lining-Deformation Co-occurrence Pattern, Structural-Hydraulic Co-occurrence Pattern, Root-Joint Co-occurrence Pattern, and Connection-Sediment Co-occurrence Pattern. Multi-method validation using five alternative weighting schemes confirmed cluster robustness, with Adjusted Rand Index values demonstrating perfect agreement within methodological families (ARI = 1.000) and moderate-to-good cross-method agreement (ARI = 0.438–0.748). Resolution parameter optimisation using Adjusted Mutual Information revealed stable clustering across threshold values 4–8 (AMI $>$ 0.79) and resolution parameters 0.8–1.2 (AMI $>$ 0.64). Further, mechanisms exhibited significant associations with pipe characteristics (diameter, age, material) and environmental context (district, land use, traffic intensity).

The framework enables three transformative improvements to asset management practice. First, mechanism-specific inspection strategies replace uniform protocols, concentrating resources on high-risk populations identified through feature associations and spatial clustering patterns. Second, defect centrality analysis identifies statistically prominent co-occurrence nodes within each mechanism; while these nodes cannot be confirmed as causal early-warning indicators without longitudinal validation, their broad statistical associations across diverse defect combinations support elevated monitoring priority as a precautionary measure. Third, rehabilitation strategies can be matched to deterioration pathways rather than aggregate condition scores, optimising intervention effectiveness and capital efficiency.

Important limitations constrain causal interpretation. The cross-sectional nature of inspection data precludes temporal validation of proposed deterioration sequences; longitudinal tracking of individual pipes through multiple inspection cycles would establish defect precedence and progression rates. Future research should integrate longitudinal analysis, incorporate operational variables, and validate mechanisms through forensic investigation of documented failures.

Despite these limitations, the demonstrated associations between mechanisms and measurable infrastructure attributes provide immediate actionable guidance for utilities. Geographic concentration of Root-Joint Co-occurrence Pattern enables spatial targeting; material-specific prevalence of Lining-Deformation Co-occurrence Pattern supports risk-based prioritisation; and the high prevalence of Structural-Hydraulic Co-occurrence Pattern justifies systematic condition-based monitoring. The framework's generalizability across different utilities, climate zones, and demographic contexts remains to be established, representing a priority for future multi-city comparative studies. This mechanism-based approach fundamentally shifts drainage asset management from reactive failure response toward proactive deterioration management grounded in physical understanding of failure processes.

CRediT authorship contribution statement

Dramani Arimiyaw: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Conceptualization. **Tarek Zayed:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Jingchao Yang:** Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Beenish Bakhtawar:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sherif Abdelkhalek:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mohamed Nashat:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the Research Grants Council (RGC)-General Research Fund (GRF) under grant number 15209022. The authors would also like to thank the Hong Kong Drainage Services Department (DSD) for the data support and the Hong Kong Utility Training Institute (UTI) for providing the *Hong Kong Conduit Condition Evaluation Codes* technical reference materials and for their valuable industry communication and guidance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tust.2026.107708.

Data availability

Data will be made available on request.

References

Abdelkhalek, S., Zayed, T., 2023. A multi-tier deterioration assessment models for sewer and stormwater pipelines in Hong Kong [Article]. J. Environ. Manage. 345, 118913. https://doi.org/10.1016/j.jenvman.2023.118913.

Alqahtani, F.K., Alsharef, A., Hommadi, G.M., Alammari, M.A., 2023. Assessment framework for the maintainability of sewer pipeline systems. Appl. Sci. 13 (21), 11828.

Arimiyaw, D., Zayed, T., Nashat, M., Yang, J., Kuoribo, E., Vimonsatit, V., Askarinejad, H., Ramezanianpour, M., Singh, A., Yazdani, S., 2025. Machine Learning Approach to root Intrusion Prediction in Urban Sewers using CCTV and Environmental Features. Proceedings of International Structural Engineering and Construction 12 (1). https://doi.org/10.14455/isec.2025.12(1).Env-03.

Balekelayi, N., Tesfamariam, S., 2020. Geoadditive quantile regression model for sewer pipes deterioration using boosting optimization algorithm [Article]. Sustainability (Switzerland) 12 (20), 1–24. https://doi.org/10.3390/su12208733. Article 8733.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Stat. Mech: Theory Exp. 2008 (10), P10008.

Carvalho, G., Amado, C., Brito, R.S., Coelho, S.T., Leitão, J.P., 2018. Analysing the importance of variables for sewer failure prediction [Article]. Urban Water J. 15 (4), 338–345. https://doi.org/10.1080/1573062X.2018.1459748.

Chughtai, F., Zayed, T., 2008. Infrastructure condition prediction models for sustainable sewer pipelines [Article]. J. Perform. Constr. Facil 22 (5), 333–341. https://doi.org/10.1061/(ASCE)0887-3828(2008)22:5(333).

Chui, S., Leung, J.K., Chu, C., 2006. The development of a comprehensive flood prevention strategy for Hong Kong. International Journal of River Basin Management 4 (1), 5–15.

Daher, S., Zayed, T., Elmasry, M., Hawari, A., 2018. Determining the relative weights of sewer pipelines' components and defects. J. Pipeline Syst. Eng. Pract. 9 (1), 04017026.

Elmasry, M., Hawari, A., Zayed, T., 2017. Defect-based deterioration model for sewer pipelines using Bayesian belief networks [Article]. Can. J. Civ. Eng. 44 (9), 675–690. https://doi.org/10.1139/cjce-2016-0592.

Falter, B., 1996. Structural analysis of sewer linings. Tunn. Undergr. Space Technol. 11, 27–41.

Fontecha, J.E., Agarwal, P., Torres, M.N., Mukherjee, S., Walteros, J.L., Rodríguez, J.P., 2021. A Two-Stage Data-Driven Spatiotemporal Analysis to Predict failure risk of Urban Sewer Systems Leveraging Machine Learning Algorithms [Article]. Risk Anal. 41 (12), 2356–2391. https://doi.org/10.1111/risa.13742.

Goode, M., Abu, J.J., Alves, P.B., Woerner, E.M., Levell-Young, T., Smith-Hams, T., Volpitta, A., Crews, R., Brown, M., Rosenberg Goldstein, R.E., 2025. A peek at leaks and basement backups: a pilot survey exploring the impacts and outcomes of untreated sewage in homes. Environ. Res. Commun. 7 (4), 045025.

Harvey, R.R., McBean, E.A., 2014a. Comparing the utility of decision trees and support vector machines when planning inspections of linear sewer infrastructure [Article]. J. Hydroinf. 16 (6), 1265–1279. https://doi.org/10.2166/hydro.2014.007.

Harvey, R.R., McBean, E.A., 2014b. Predicting the structural condition of individual sanitary sewer pipes with random forests [Article]. Can. J. Civ. Eng. 41 (4), 294–303. https://doi.org/10.1139/cjce-2013-0431.

Keung, K. L., Lee, C. K. M., Ng, K., & Yeung, C.-K. (2018). Smart city application and analysis: Real-time urban drainage monitoring by iot sensors: A case study of Hong Kong. 2018 IEEE international conference on industrial engineering and engineering management (IEEM).

Khazraeializadeh, S., Gay, L.F., Bayat, A., 2014. Comparative analysis of sewer physical condition grading protocols for the City of Edmonton [Article]. Can. J. Civ. Eng. 41 (9), 811–818. https://doi.org/10.1139/cjce-2014-0077.

Kuliczkowska, E., Parka, A., 2017. Management of the risk of tree and shrub root intrusion into sewers. Urban For. Urban Green. 21, 1–10.

Kumar, S., Deshpande, V., Agarwal, M., Rathnayake, U., 2024. Forecasting particle Froude number in non-deposition scenarios within sewer pipes through hybrid machine learning approaches [Article]. Results Eng. 22, 102320. https://doi.org/10.1016/j.rineng.2024.102320.

Law, T.M., Moore, I.D., 2007. Numerical modeling of tight fitting flexible liner in damaged sewer under earth loads. Tunn. Undergr. Space Technol. 22 (5–6), 655–665.

Li, X., Khademi, F., Liu, Y., Akbari, M., Wang, C., Bond, P.L., Keller, J., Jiang, G., 2019. Evaluation of data-driven models for predicting the service life of concrete sewer pipes subjected to corrosion [Article]. J. Environ. Manage. 234, 431–439. https://doi.org/10.1016/j.jenvman.2018.12.098.

Loganathan, K., Najafi, M., Kermanshachi, S., Maduri, P.K., Pamidimukkala, A., 2024. Inspection prioritization of gravity sanitary sewer systems using supervised machine learning algorithms [Article]. Article 9 Journal of Infrastructure Preservation and Resilience 5 (1). https://doi.org/10.1186/s43065-024-00101-3.

Lubini, A.T., Fuamba, M., 2011. Modelling of the deterioration timeline of sewer systems [Article]. Can. J. Civ. Eng. 38 (12), 1381–1390. https://doi.org/10.1139/L11-103.

Mashford, J., Marlow, D., Tran, D., May, R., 2011. Prediction of sewer condition grade using support vector machines [Article]. J. Comput. Civ. Eng. 25 (4), 283–290. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000089.

Nashat, M., Zayed, T., Yang, J., Arimiyaw, D., 2025. Comprehensive insights into sewer corrosion: interlinked factors, prediction models, and mitigation approaches. Front. Environ. Sci. Eng. 19 (12), 159.

Newman, M.E., Girvan, M., 2004. Finding and evaluating community structure in networks. Phys. Rev. E 69 (2), 026113.

Nguyen, L.V., Razak, S., 2023. Predicting sewer structural condition using hybrid machine learning algorithms [Article]. Urban Water J. 20 (7), 882–896. https://doi.org/10.1080/1573062X.2023.2217430.

Nguyen, L.V., Seidu, R., 2022. Application of Regression-based Machine Learning Algorithms in Sewer Condition Assessment for Ålesund City, Norway [Article]. Article 3993 Water (switzerland) 14 (24). https://doi.org/10.3390/w14243993.

Nguyen, L.V., Bui, D.T., Seidu, R., 2022. Comparison of Machine Learning Techniques for Condition Assessment of Sewer Network [Article]. IEEE Access 10, 124238–124258. https://doi.org/10.1109/ACCESS.2022.3222823.

Que, X., Checconi, F., Petrini, F., & Gunnels, J. A. (2015). Scalable community detection with the louvain algorithm. 2015 IEEE international parallel and distributed processing symposium.

Randrup, T.B., McPherson, E.G., Costello, L.R., 2001. Tree root intrusion in sewer systems: review of extent and costs. J. Infrastruct. Syst. 7 (1), 26–31.

Sousa, V., Matos, J.P., Matias, N., 2014. Evaluation of artificial intelligence tool performance and uncertainty for predicting sewer structural condition [Article]. Autom. Constr. 44, 84–91. https://doi.org/10.1016/j.autcon.2014.04.004.

Spasojevic, A., Mair, R., Gumbel, J., 2007. Centrifuge modelling of the effects of soil loading on flexible sewer liners. Géotechnique 57 (4), 331–341.

Vipulanandan, C., & Liu, J. (2005). Sewer pipe-joint infiltration test protocol developed by CIGMAT. In *Pipelines 2005: Optimising Pipeline Design, Operations, and Maintenance in Today's Economy* (pp. 553-563).

Woldesellasse, H., Tesfamariam, S., 2024. Data augmentation using conditional generative adversarial network (cGAN): applications for sewer condition classification and testing using different machine learning techniques [Article]. J. Hydroinf. 26 (7), 1471–1489. https://doi.org/10.2166/hydro.2024.135.

Yang, J., Arimiyaw, D., Zayed, T., Nashat, M., Liu, X., Ibrahim, A., 2025. Survival analysis framework for sewer failure time: evidence from Hong Kong. npj Clean Water 8 (1), 91.

Yang, J., Zayed, T., Arimiyaw, D., Nashat, M., Taiwo, R., Alfalah, G., Liu, X., Ibrahim, A., 2025a. When population science meets urban sewer networks: Decoding remaining life using life table analytics. Water Res. X 100467.

Yang, J., Zayed, T., Arimiyaw, D., Nashat, M., Liu, X., Ibrahim, A., 2025. A systematic review of service reliability in urban sewer systems: from failure analysis to preventive management. Reliab. Eng. Syst. Safe. 112025.

Yang, J., Zayed, T., Arimiyaw, D., Xiao, R., 2025. A comprehensive review of influential factors and predictive techniques of time to failure for sewer pipes. Tunnell. Undergr. Space Technol. 157, 106357.

Yang, J., Zayed, T., Liu, X., Arimiyaw, D., Nashat, M., & Ibrahim, A. (2025). Towards predictive modeling of time to sewer pipe failure: a preliminary exploratory study combining statistical analysis and AI techniques. Doi: 10.22260/CRC-CSCE-2025/0014.

Yang, J., Zayed, T., Arimiyaw, D., Nashat, M., Liu, X., Ibrahim, A., 2026. Structural failure in urban sewer pipelines: a comprehensive review from analysis to intervention techniques. Tunn. Undergr. Space Technol. 167, 107027.

Yin, X., Chen, Y., Bouferguene, A., Al-Hussein, M., 2020. Data-driven bi-level sewer pipe deterioration model: Design and analysis [Article]. Autom. Constr. 116, 103181. https://doi.org/10.1016/j.autcon.2020.103181.

Yusuf, W., Alaka, H., Ahmad, M., Ebenezer, W., Ajayi, S., Toriola-Coker, L. O., & Ahmed, A. (2022). Automatic surface encrustation detection in concrete sewers using CCTV inspection footage: A fusion of deep learning and big data. Available at SSRN 4331293.

Zamanian, S., Shafieezadeh, A., 2023. Age-dependent failure probabilities of corroding concrete sewer pipes under traffic loads [Article]. Structures 52, 524–535. https://doi.org/10.1016/j.istruc.2023.03.132.