ENVIRONMENTAL RESEARCH
LETTERS

PURPOSE-LED PUBLISHING™

LETTER • OPEN ACCESS

Enhancing assessment of direct and indirect exposure of settlement-transportation systems to mass movements by intergraph representation learning

To cite this article: Joshua Dimasaka *et al* 2024 *Environ. Res. Lett.* **19** 114055

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

Enhancing assessment of direct and indirect exposure of settlement-transportation systems to mass movements by intergraph representation learning

Joshua Dimasaka[1,5,*], **Sivasakthy Selvakumaran**[2] **and Andrea Marinoni**[3,4]

[1] Department of Earth Sciences, University of Cambridge, Cambridge, United Kingdom
[2] Department of Engineering, University of Cambridge, Cambridge, United Kingdom
[3] Department of Physics and Technology, UiT the Arctic University of Norway, Tromsø, Norway
[4] Department of Computer Science and Technology, University of Cambridge, Cambridge, United Kingdom
[5] Present address: Cambridge University Centre for Risk in the Built Environment, Department of Architecture, University of Cambridge, Cambridge, United Kingdom.
[*] Author to whom any correspondence should be addressed.

E-mail: jtd33@cam.ac.uk

Keywords: mass movements, exposure, susceptibility, settlement, road network, graph-based machine learning, clustering

Supplementary material for this article is available online

OPEN ACCESS

RECEIVED
10 July 2024

REVISED
11 September 2024

ACCEPTED FOR PUBLICATION
30 September 2024

PUBLISHED
14 October 2024

Abstract

Amidst the intensifying extreme rainfall patterns due to climate change, global early warning systems for mass movements (e.g. landslides, avalanches) need to provide not only the coarsely aggregated danger reports, but also the necessary fine details to understand its potential implications on critical infrastructures such as transportation systems. In this study, we introduce a novel 'intergraph' method that enhances the exposure information using a graph-based machine learning implementation on the hydrological and geological characteristics of mass movements and the underlying connectivity of settlement-transportation systems. Demonstrating the entire country of Norway and the 2020 Gjerdrum quick clay incident as a case study, we integrated the assessment of both direct and indirect exposure information of settlement-transportation systems and their daily 1 km-by-1 km susceptibility map, which were derived from the 68 934 mass movement incidents since 1957 and the connectivity information of 4778 settlements and 257 000 km road networks. Our findings achieved 86.25% accuracy, providing a distribution of improved susceptibility estimates and identifying critical settlements in near-real-time. By interacting the graphical representations of the shared causal drivers of susceptibility and the settlement-transportation system connectivity, our study extends our understanding of the exposure of multiple interacting settlements with a high granularity degree in a unified approach.

1. Introduction

Over the years, amidst the increasing extremes of rainfall and snow that trigger landslides and avalanches [1, 2], the United Nations Office for Disaster Risk Reduction has underscored the global need for the development of accurate, reliable, and localized early warning systems to effectively reduce and mitigate disaster risks [3]. For example, in the aftermath of the widespread 2020 Gjerdrum quick clay landslide that displaced over 1600 residents and killed 10 in Norway [4–6], many Norwegian parliament members stressed the need to improve the quality of risk information to such mass movements [7]. As Norway has experienced over 84 500 reported incidents such as avalanches, rockslides, and slippage since 1900 [8, 9], the understanding on cascading, compounding, and indirect effects has also gained importance, along with the direct effects of mass movements to every community [10].

Current state-of-the-art efforts of the Norwegian early warning system use a matrix-based approach

to combine four classes of catchment susceptibility and daily rainfall intensity to determine warning levels (low, moderate, high, very high), which can be very sensitive and dependent on the predefined categorization [11, 12]. The outputs are deterministic with only a single category assigned to the county or smaller villages, which does not show a disaggregated distribution of the warning levels along road networks or within the vicinity of settlements [13], potentially leading to a poor perception of risk and costly local mitigation measures.

To address these challenges, sophisticated models and high-resolution mapping have recently been introduced at the intersection of graph-based machine learning and remote sensing as they combine the nonlinearity and graphical information of various geospatial datasets (e.g. spatial, causal, or any temporal connections) [14, 15]. To this end, we introduce the 'intergraph', a novel approach that extends the exposure information by explicitly capturing the underlying connectivity of settlement-transportation systems and their graphical representations on the map. It transforms the information of the hydrological and geological characteristics of 68 934 mass movements incidents and the 4778 formal settlements with 257 000 km road networks into graphical representations (i.e. nodes connected by edges). Using graph-based machine learning, it generates a daily 1 km-by-1 km susceptibility map and quantifies the direct and indirect exposure of all settlements to mass movements in near-real-time.

2. Related work

Prior studies on susceptibility mapping have underscored various modelling challenges such as the highly conservative estimates [16, 17] and the spatial correlation [18] due to regional climatic differences [19, 20]. Early efforts in applying statistical and machine-learning techniques [18, 21–23] have achieved satisfactory results. However, each of these methods show significantly different performance because of their respective inherent uncertainties. Among these several techniques, the ensemble approach, which could handle different model uncertainties, has gained the highest precision in combination with decision trees [24], neural networks [25], logistic regression [26], and logistic model trees [27]. However, these still have limited explainability as they disregarded spatial correlation and region-specific attributes.

Recent studies on integrating graphical information with machine learning such as graph neural networks [28–30] have allowed several applications in disaster management [31], air pollutant estimation [32], transportation disruption detection [33], and early warning system [34, 35], as they capture the spatial connectivity and the attribute similarity of every data point. Despite these potential improvements, no studies have also been done to systematically integrate the insights from susceptibility maps with exposure patterns such as population, settlements, and roads, thereby limiting the current early warning system [13] at the county-level granularity with predefined large polygons that excessively aggregate the information on the map.

In this study, we, therefore, combined the ensemble approach and graph neural networks to generate a daily susceptibility map of Norway. We then used the graphical information of its derived susceptibility values as attributed characteristics of the road connectivity and the vicinity of settlement-transportation systems, thereby evaluating the degree of exposure of every populated settlement connected by roads.

3. Methodology

In this section, we present a detailed summary of the major procedures from the data preparation to the implementation of our proposed 'intergraph' representation learning, which consists of two main steps: *Supervised Ensemble Graph Neural Network* for susceptibility mapping and *Unsupervised Spectral Graph Clustering* for settlement exposure evaluation.

3.1. Data preparation

3.1.1. Labels and features.

We used 68 934 incidents of mass movements (e.g. rockslide, landslide, avalanche, slippage, mudslide, flood landslide, and unspecified) reported since 1957 [36]. Their geographical and temporal information guided the extraction of relevant hydrological and geological characteristics. As shown in table 1, we used two types of publicly available geospatial features: static and date-specific dynamic. See section S1 in the supplementary material for more details on handling the large file sizes of these datasets.

3.1.2. Grid discretization

To facilitate a consistent representation of features for the susceptibility mapping, we discretized all geospatial features using a 1 km-by-1 km grid, resulting in a map with a 1195×1550-array. For the settlement exposure evaluation to realistically capture the detail of road networks, we increased the detail of discretization from 1 km to 50 m (i.e. 20 times finer), resulting in a higher data storage requirement and longer processing time. To resolve this, we employed a county-level masking to subdivide the large areal extent of Norway. See section S2 in the supplementary material for more details on the geographical coverage of each mask.

Table 1. Geospatial features.

Type	Details
Static	steepness [37]
Static	susceptibility category [38, 39]
Static	shallow subsurface lithology class[a] [40]
Static	land cover class[a] [41]
Static	slope angle class[a] [42]
Dynamic	total rainfall[b] [43]
Dynamic	mean temperature[b] [43]
Dynamic	snow depth[b] [44]
Dynamic	snow water equivalent[b] [44]
Dynamic	fresh snow water equivalent[b] [44]

[a] These are ancillary maps of the European Landslide Susceptibility Map (ELSUS v2) [38, 39] from the European Soil Data Centre [45–47].

[b] These are accessed from the Norwegian Meteorological Institute [48].

3.1.3. Ensemble and dataset split

Out of the 508 182 grid points that cover the entire map of Norway, only about 16 000 unique grid points represent the locations of the past 68 934 incidents. Given the limited computational capacity, we set 500 as the number of randomly sampled unique grid points with incident labels, resulting in 32 subdatasets (i.e. $16\,000/500 = 32$). Then, for each unique grid point or location, we randomly sampled a representative past incident. For every subdataset of 500 grid points of past incidents, we included 500 grid points of non-incident that were sampled at locations without reported or unknown incidents.

We split each subdataset using 70%, 15%, and 15%, for training, validating, and testing, respectively, to avoid model overfitting. As a result, we prepared 32 subdatasets wherein each subdataset is comprised of 1000 unique grid points of equally sampled incidents and non-incidents. In the following sections, we trained each subdataset using a machine-learning model called graph convolutional neural network (GCN) [29], which effectively results in an ensemble of 32 GCN models.

3.1.4. Formal settlements and road networks

To understand how the susceptibility values may affect the exposure patterns, we used the available datasets of urban formal settlements (i.e. a minimum population of 200 and a maximum inter-house distance of 50 m) [49] and road networks [50], respectively updated in 2022 and 2020. These are represented as geospatial vector polygons and polylines, respectively.

3.2. Intergraph representation learning

As shown in figure 1, the 'intergraph' models the interaction of two graphical representations wherein the information of the node outputs from the first method (*Supervised Ensemble Graph Neural Network*) is used as an indicator of the strength of the edge connection between the nodes in the graphical representation of the second method (*Unsupervised Spectral Graph Clustering*).

3.2.1. Supervised ensemble graph neural network

As shown in figure 2, from the pool of 68 934 incidents, we started with identifying the dates of a randomly selected set of incident labels and extracting their associated date-specific features. With knowledge on the location of these labeled points, we also randomly sampled an equal number of non-incident points from locations with no reported or unknown incidents.

The combined random samples of incident and non-incident points formed a resulting subdataset S_k, which is further represented using two graphical structures: attribute-aware G_k^A (i.e. the similarity of geospatial features using cosine similarity metric); and neighborhood-aware G_k^N (i.e. the degree of proximity using the relative geographical distances). We combined these two graphical structures into a single weighted G_k after investigating a range of possible ratio values. In section S3, we present the supplementary details on this ratio and other optimal parameters, as we trained the 32 GCN models multiple times.

For each subdataset S_k, we trained its corresponding GCN model M_k to optimize its parameters. In this way, when the GCN model tries to predict the probability, these incidents have high probability assigned. The graphical information of G_k provided the training with an embedded inductive structure for learning as to how the incremental updates are propagated with respect to the degree of similarity and relationship among the selected points in a given S_k. We trained each GCN model M_k until the validation loss reached a minimum. In section S4, we present a supplementary mathematical formulation of this implementation.

To analyze the predictive performance of the ensemble, we generated the receiver operating characteristic (ROC) curve to observe how true positive rates (*TPR*) change with false positive rates (*FPR*). A desirable area under the ROC curve (AUC) should be very close to 1.0, wherein 0.5 indicates random guessing only while $(0.7, 0.8]$ is acceptable, $(0.8, 0.9]$ is excellent, and $(0.9, 1.0]$ is outstanding [51].

As each GCN model M_k of the ensemble predicts the susceptibility on the map, we can aggregate their estimates and present their differences or variation. In this way, any percentiles can also be used to specify a range of reasonable estimates of susceptibility due to the model uncertainty.

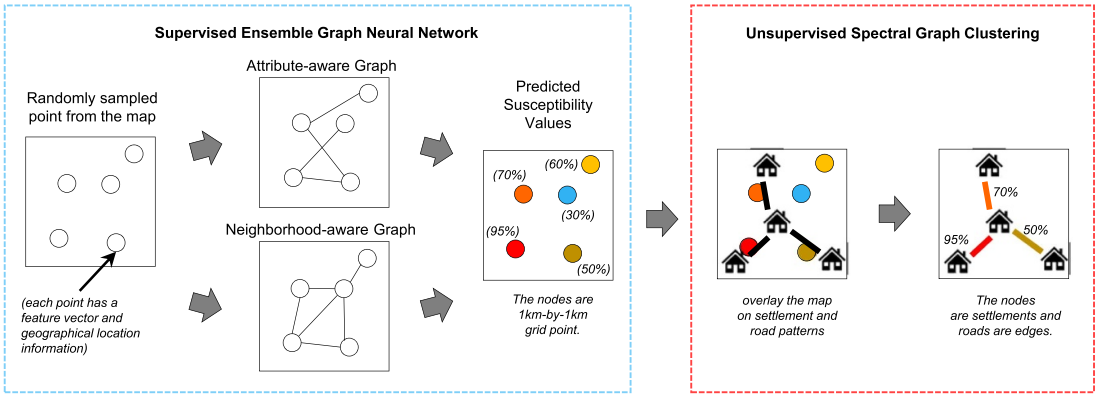

Figure 1. An overview of the 'intergraph' implementation that consecutively illustrates the random sampling of points or nodes, establishing graphical relationships, predicting susceptibility values, constructing the graphical structure between settlements and roads, assigning the edge weight from the predicted susceptibility values, and clustering the resulting settlement network. The attribute-aware and neighborhood-aware graphs pertain to the degree of relationships between nodes in terms of the similarity of geospatial features and the relative geographical distance, respectively.

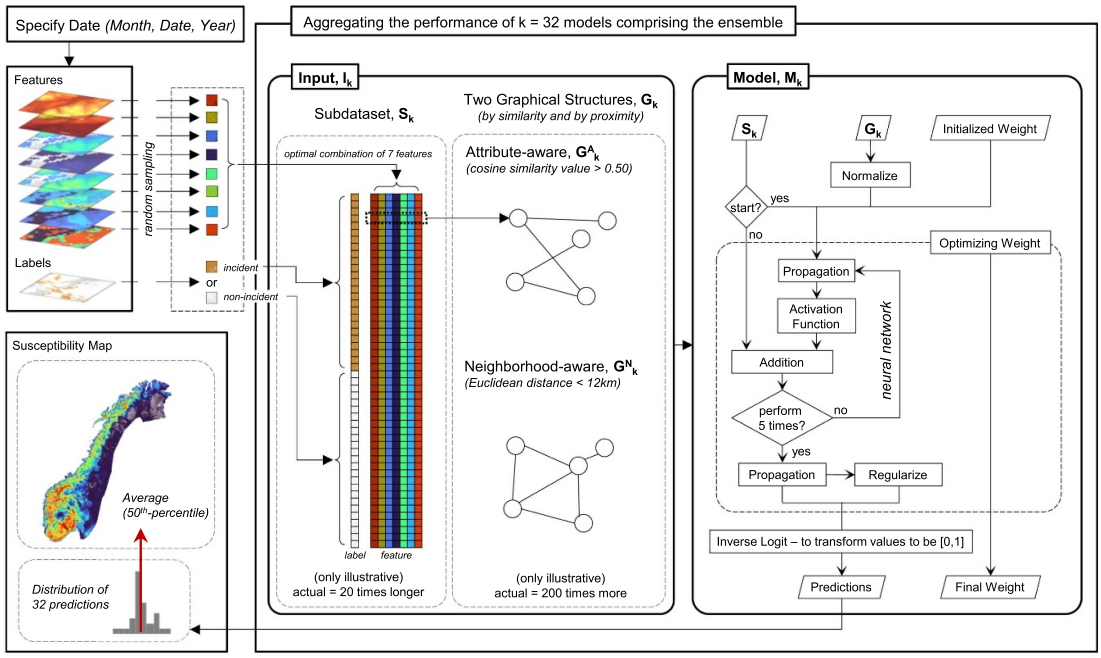

Figure 2. A schematic flowchart of the supervised ensemble graph neural network.

3.2.2. Unsupervised spectral graph clustering

As shown in figure 3, we used a county-level two-step graph structuring approach to model the connections between formal settlements and roads. The first graphical structure treats every pixel or grid point of formal settlements and road networks on the map as nodes of a simplicial complex [52, 53]. This means that a connection exists if two pixels or points are adjacent to each other by one step in all directions. This simplicial complex generally describes the topology of roads and settlements, which has enabled us to extract other information such as the shortest path between any two pixels.

Using the length and profile of extracted shortest paths that could indicate a probable edge between any two settlements as nodes, we developed the second graphical structure wherein (1) a random point within the vicinity of a particular settlement represents the node; and (2) a characteristic value (e.g. the maximum susceptibility value along the shortest path between any two settlements) serving as the relative weight of their connection.

We then performed an iterative spectral graph clustering wherein we observed which settlement gets isolated when we increase the susceptibility cutoff value from 0 to 1. This involves the removal of an

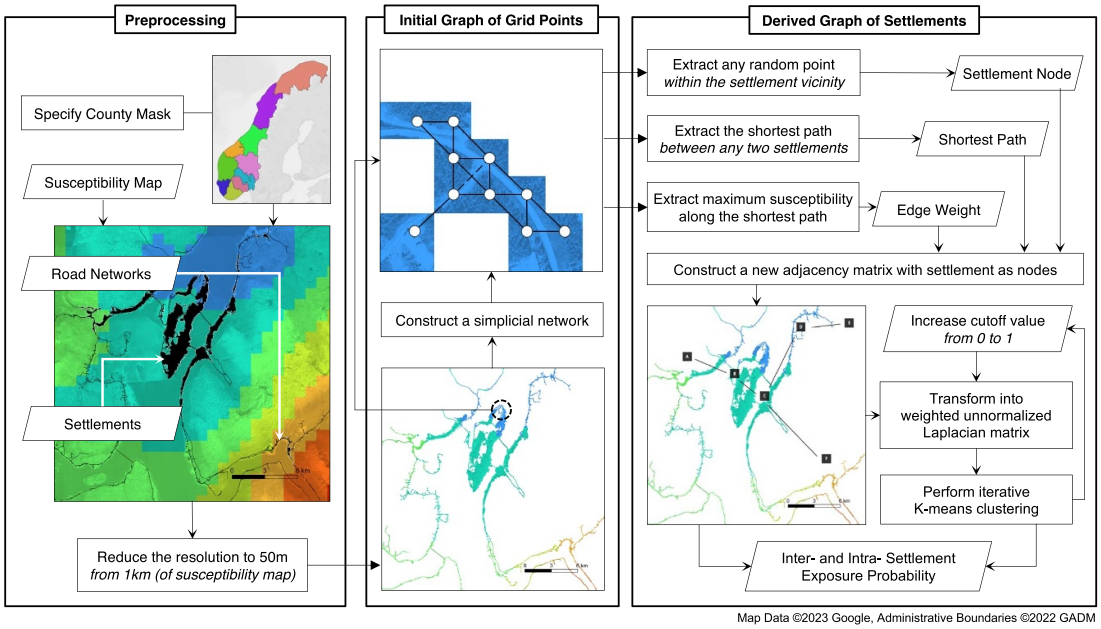

Figure 3. A schematic flowchart of the unsupervised spectral graph clustering.

edge connection if its assigned characteristic value is less than or equal to the cutoff value. For every iteration on the modified graphical structure, we implemented K-means clustering using the spectrum (i.e. eigenvectors) of the unnormalized graph Laplacian transformation of our derived graphical representation of settlements [54]. We also used its information on eigenvectors that correspond to zero eigenvalues because it indicates the number of connected sub-components of our original graph representation [55].

As a result, we extracted the lowest cutoff value that would cause an isolation of a settlement node. We defined this value as the *indirect (or inter)* exposure probability because it signifies the minimum susceptibility of roads to trigger an isolation among the connected settlements. We also presented the *direct (or intra)* exposure probability that describes the aggregated susceptibility value of an area to mass movements within its boundaries. In section S5, we present a supplementary pseudocode with the steps in deriving these indirect and direct exposure probabilities.

4. Results and discussion

4.1. Mapping mass-movement susceptibility

In figure 4, we present the overall performance of our ensemble for mapping the susceptibility to mass movements. For all the 15%-testing portions of 32 subdatasets combined together, the resulting ROC curve achieved an accuracy of 86.25% AUC, effectively assigning higher probability values to areas that are susceptible to mass movements, as compared to the random-guessing line. This indicates that the ensemble can accurately predict the absence and presence of a mass movement incident with 75% and 83% success rates, respectively. These success rates were derived from the optimal operating point of the ROC curve, which is equivalent to the minimum misclassification costs (i.e. a tradeoff between *TPR* and *FPR*). In section S6, we present a supplementary discussion on this optimal operating point.

Table 2. Feature selection frequency in the ensemble.

Type	Details	Frequency (↑)
Static	steepness	32/32 (all)
Static	susceptibility category	32/32 (all)
Static	slope angle class	32/32 (all)
Dynamic	mean temperature	32/32 (all)
Dynamic	total rainfall	29/32
Static	land cover class	22/32
Dynamic	snow depth	13/32
Dynamic	snow water equivalent	12/32
Static	shallow subsurface lithology class	11/32
Dynamic	fresh snow water equivalent	9/32

Considering the individual GCN models of the ensemble, we summarized the frequency of every selected feature in table 2 to understand the degree of relevance of each geospatial feature layers. All 32 GCN models consistently used static steepness, static susceptibility category, static slope angle class, and dynamic mean temperature. This is closely followed by the dynamic total rainfall and static land cover class. This confirms

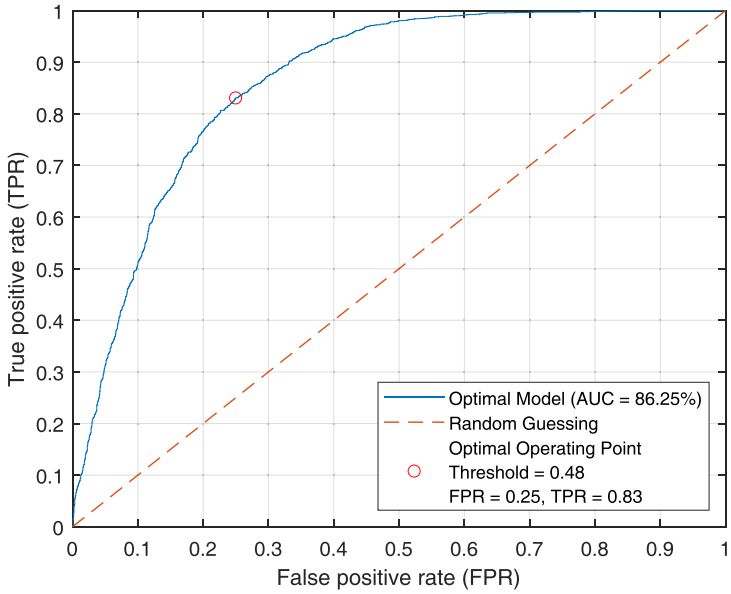

Figure 4. ROC curve of the trained ensemble. As the curve comprises different pairs of points of *FPR* and *TPR*, minimizing the misclassification costs yields 0.48 as the corresponding optimal operating point, which becomes the new relative midpoint for the resulting susceptibility values on the map.

that the predictive performance of our resulting ensemble depends on these well-known physical triggers and geological characteristics. This also implies that predictive capability can be further improved using the near-real-time information on mean temperature and total rainfall, instead of using the static susceptibility category alone.

Moreover, we observed that the less frequently selected features were dynamic snow depth, dynamic snow water equivalent, static shallow subsurface lithology class, and dynamic fresh snow water equivalent. The relatively weaker contribution of dynamic snow features may be attributed to its passive characterization of the snow volume and not indicative of snow movement. This case also highlights the influence of other feature layers such as the triggering effects of dynamic total rainfall and static steepness. The limited number of classes of shallow subsurface lithology also explains its weaker relevance on the majority of GCN models. Nevertheless, the resulting ensemble has shown that the graphical information of these geospatial layers enabled the encoding of causal inductive structures with several static and dynamically changing features.

To demonstrate the effectiveness of the proposed approach in improving our estimates of susceptibility to mass movements, we considered the 2020 Gjerdrum quick clay landslide as a case study [4], as shown in figure 5. In that case, figure 5(m) (red circle) shows that the the current Norwegian early warning system reported no significant danger for avalanches [56] and a very low probability for landslides [57] for the municipality of Gjerdrum. In contrast, figure 5(k) (white circle) shows a significantly higher prediction of susceptibility estimates at 58% with a standard deviation of 25%. This generally agrees with the occurrence of the said incident and the large rainfall patterns (46 mm) in figure 5(a). Considering the special characteristics of the Gjerdrum event [4], this result underscores how well the proposed approach is able to generalize the analysis of the susceptibility estimation to a wide range of mass movements events whilst ensuring the robustness of the outcomes because of their shared causal drivers and offering high-resolution maps compared to the county-level presentation of the current system.

Combining the high generalization performance of the proposed approach together with the intrinsic flexibility of the data representation provided by the graph structure allows us to deeply investigate the direct and indirect effects of mass movements with a high granularity degree in the succeeding section. This is important as it presents an opportunity for early warning systems based on the proposed methodology to extend their current capacity, hence incorporating socioeconomic factors with environmental data analysis in a single data architecture.

4.2. Evaluating settlement exposure

Both direct (*intra)* and indirect (*inter)* analyses are equally important in understanding the total exposure characteristics of a settlement. For instance, even though a settlement (e.g. danger zones but well-connected to roads) may have a low minimum probability to trigger inter-settlement isolation, it may still have a high intra-settlement exposure probability because it may be situated next to an

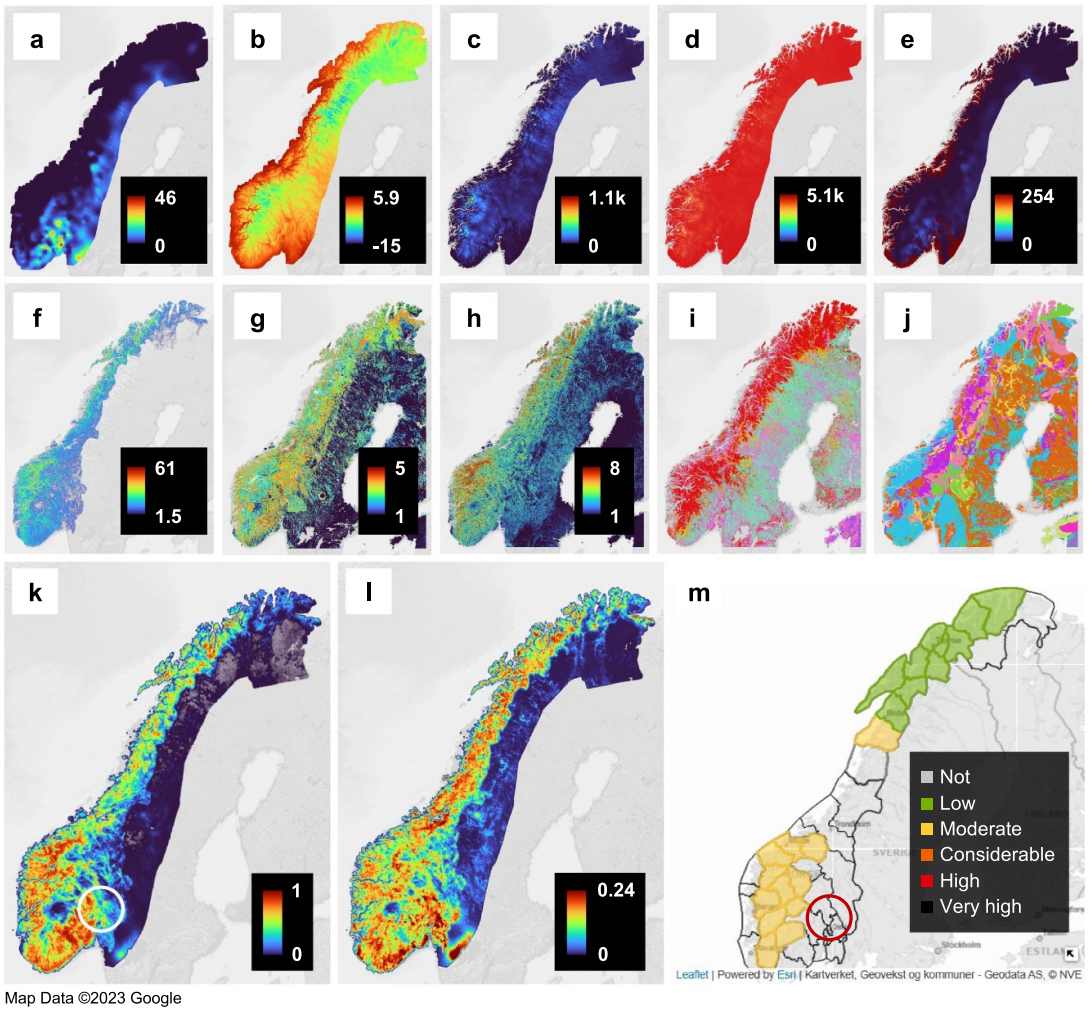

Figure 5. The dynamic feature maps for 30 December 2020: (a) total rainfall (mm d^{-1}), (b) mean temperature (Celsius), (c) snow depth (cm d^{-1}), (d) snow water equivalent (mm d^{-1}), and (e) fresh snow water equivalent (mm d^{-1}). Static feature maps of (f) steepness (degrees), (g) ELSUS susceptibility (category), (h) slope angle class (category), (i) land cover class (category), and (j) lithology class (category). The corresponding resulting maps for 30 December 2020, when the Gjerdrum quick clay landslide incident occurred: (k) average susceptibility map and (l) calculated standard deviation using the estimates of 32 GCN models. (m) Actual forecast across Norway for mass movement alert with a red circle showing the location of the quick clay incident (corresponding to the white circle in figure 5(k).

eroding soil mass with imminent danger within its vicinity.

As an illustrative example following our results for 30 December 2020, figure 6 shows that the relatively remote village of Leinesfjord has high inter-settlement exposure probability (69%) but low intra-settlement exposure probability (21%) whereas the opposite case occurred for the village of Brattvåg, which is connected to many neighboring villages. This evaluation of exposure probability for roads to trigger inter-settlement isolation uniquely presents a new perspective by capturing the effect of meaningful interaction of settlements. See section 7 in the supplementary material for a summary of these two probabilities with population information in the counties of Oslo and Viken, where the significant rainfall pattern was mostly observed.

Enabling a computationally efficient large-scale evaluation of settlement exposure, our county-level two-step graph structuring approach has also established a flexible framework that can incorporate valuable considerations such as (1) the inter-county connectivity, which can provide additional information on how a settlement may access mitigation assistance from neighboring settlements from different counties; (2) the use of multiple shortest paths, which can investigate all probable connections for cooperation between any two settlements; and (3) the variation of physical vulnerability along each connecting road, which can accurately implement a targeted approach in prioritizing critical retrofit measures for transportation networks. Consequently, this offers a future advancement and extension of the current state-of-the-art approach, which would aid the decision-making process of Norwegian county governors in charge of coordinating the mitigation efforts of various municipalities within their respective county-level jurisdictions.

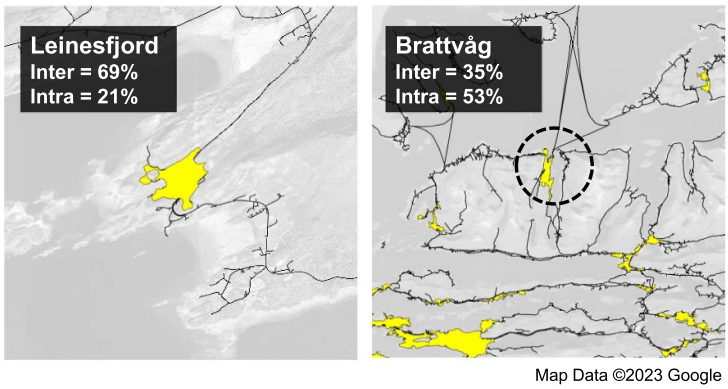

Figure 6. A comparison between the villages (*yellow polygons*) of Leinesfjord (*left*) and Brattvåg (*right*).

5. Conclusion and future work

Our proposed novel 'intergraph' representation learning has opened an opportunity to develop solutions for climate change adaptation and mitigation by understanding the graphical relationships of different geospatial datasets of susceptibility and exposure patterns of settlement-transportation systems. Our study has specifically developed a proof of concept that is computationally efficient for the future advancement of the automated Norwegian and other global early warning system from mass movements, which could effectively identify the high-risk settlements as pertinent and urgent information for policymakers and regional county governors. For future work, we recommend extending this study with diverse climate scenarios to understand the long-term impacts of mass movements and their cascading effects to the communities. We also recommend investigating the usefulness of other near-real-time geospatial information derived from satellite imagery and other graph representation learning techniques.

Data availability statement

The complete set of data (40GB) and code is available in our publicly available Zenodo repository (https://zenodo.org/record/8099812) [58]. The code and its documentation are available in our GitHub repository (https://github.com/IMPACTSquad/InterGraphNorwayMM).

Acknowledgments

This work is funded by the UKRI Centre for Doctoral Training in Application of Artificial Intelligence to the study of Environmental Risks (EP/S022961/1), Centre for Integrated Remote Sensing and Forecasting for Arctic Operations and the Research Council of Norway (237906), the Visual Intelligence Centre for Research-based Innovation funded by the Research Council of Norway (309439), and the IMPETUS project funded by the European Union Horizon 2020 research and innovation program (101037084). We also thank the support from the AI4ER community and the Integrated Multimodal Pattern Analysis for Climate Transformation (IMPACT) research group.

ORCID iDs

Joshua Dimasaka https://orcid.org/0000-0001-9618-9431
Sivasakthy Selvakumaran https://orcid.org/0000-0002-8591-0702
Andrea Marinoni https://orcid.org/0000-0001-6789-0915

References

[1] IPCC 2023 *Weather and Climate Extreme Events in a Changing Climate* (Cambridge University Press) pp 1513–766
[2] Hanssen-Bauer I *et al* 2009 Climate in Norway 2100 *Background Information to NOU Climate Adaptation (In Norwegian: Klima i Norge 2100. Bakgrunnsmateriale til NOU Klimatilplassing)* (Norsk klimasenter)
[3] United Nations Office for Disaster Risk Reduction 2023 Summary of the high-level meeting of the united nations general assembly on the midterm review of the implementation of the sendai framework for disaster risk reduction 2015-2030 (available at: https://sendaiframework-mtr.undrr.org/media/88350) (Accessed 6 March 2024)
[4] Petley D 2023 The causes of the 30 December 2020 Gjerdrum landslide in Norway (available at: https://blogs.agu.org/landslideblog/2023/06/09/gjerdrum-landslide-report/) (Accessed 15 June 2024)
[5] L'Heureux J-S, Kristian Bache B, Hov S and Lacasse S 2022 Mitigation measures following landslides in sensitive clays, examples from Norway *Geohazards VIII*
[6] Penna I and Solberg I-L 2021 Landscape changes and bedrock reconstruction in gjerdrum area. methodological approach and main results *Technical Report* 394800 (Norges Geologiske Undersøkelse, Fence Room, Creek)
[7] Stortinget 2021 Written question from Une Bastholm (MDG) to the Minister of Petroleum and Energy (available at: www.stortinget.no/no/Saker-og-publikasjoner/Sporsmal/Skriftlige-sporsmal-og-svar/Skriftlig-sporsmal/?qid = 82767) (Accessed 20 June 2023)
[8] Laute K and Beylich A A 2018 Potential effects of climate change on future snow avalanche activity in western Norway

deduced from meteorological data *Geogr. Ann.* A **100** 163–84

[9] Jaedicke C, Lied K and Kronholm K 2009 Integrated database for rapid mass movements in Norway *Nat. Hazards Earth Syst. Sci.* **9** 469–79
[10] UNDRR & WMO 2023 Global status of multi-hazard early warning systems (available at: https://wmo.int/publication-series/global-status-of-multi-hazard-early-warning-systems-2023) (Accessed 13 March 2024)
[11] Devoli G, Bell R, and Cepeda J 2019 Susceptibility map at catchment level, to be used in landslide forecasting, Norway *Technical Report* (Norwegian Water Resources and Energy Directorate)
[12] Barfod E, Müller K, Saloranta T, Andersen J, Kristian Orthe N, Wartianien A, Humstad T, Myrabø S and Engeset R 2013 The expert tool XGEO and its applications in the Norwegian Avalanche forecasting service *Int. Snow Science Workshop Grenoble* pp 7–11
[13] Engeset R V 2013 National avalanche warning service for Norway–established 2013 *Proc. ISSW* pp 301–10
[14] Zhang X, Zhou Y'nan and Luo J 2022 Deep learning for processing and analysis of remote sensing big data: a technical review *Big Earth Data* **6** 527–60
[15] Khlifi M K, Boulila W and Riadh Farah I 2023 Graph-based deep learning techniques for remote sensing applications: techniques, taxonomy and applications - a comprehensive review *Comput. Sci. Rev.* **50** 100596
[16] Derron M H and Sletten K 2016 Method for the susceptibility mapping of snow avalanches in Norway *Technical Report* (Norwegian Water Resources and Energy Directorate)
[17] Solheim A, Sverdrup-Thygeson K and Kalsnes B 2023 Hazard and risk assessment for early phase road planning in Norway *Nat. Hazards* **119** 943–63
[18] Erener A and Düzgün H S B 2010 Improvement of statistical landslide susceptibility mapping by using spatial and global regression methods in the case of more and romsdal (Norway) *Landslides* **7** 55–68
[19] Jaedicke C *et al* 2008 Spatial and temporal variations of norwegian geohazards in a changing climate, the geoextreme project *Nat. Hazards Earth Syst. Sci.* **8** 893–904
[20] Fischer L, Rubensdotter L, Sletten K, Stalsberg K, Melchiorre C, Horton P and Jaboyedoff M 2012 Debris flow modeling for susceptibility mapping at regional to national scale in Norway *Proc. 11th Int. and 2nd North American Symp. on Landslides* pp 3–8
[21] Luo H, Liu Z, Pan Y and Rocchi I 2023 Gis-based rainfall-induced landslide susceptibility mapping: a comparative analysis of machine learning algorithms and a numerical method in kvam, Norway *EGU General Assembly Conf. Abstracts* p EGU–17051
[22] Ullah I, Aslam B, Hassan Iqbal Ahmad Shah S, Tariq A, Qin S, Majeed M and Havenith H-B 2022 An integrated approach of machine learning, remote sensing and gis data for the landslide susceptibility mapping *Land* **11** 1265
[23] ThaiPham B, Shirzadi A, Tien Bui D, Prakash I and Dholakia M B 2018 A hybrid machine learning ensemble approach based on a radial basis function neural network and rotation forest for landslide susceptibility modeling: a case study in the himalayan area, india *Int. J. Sediment Res.* **33** 157–70
[24] Merghadi A, Yunus A P, Dou J, Whiteley J, ThaiPham B, Tien Bui D, Avtar R and Abderrahmane B 2020 Machine learning methods for landslide susceptibility studies: a comparative overview of algorithm performance *Earth-Sci. Rev.* **207** 103225
[25] ThaiPham B, Bui D T, Prakash I and Dholakia M B 2017 Hybrid integration of multilayer perceptron neural networks and machine learning ensembles for landslide susceptibility assessment at himalayan area (India) using gis *Catena* **149** 52–63
[26] Chen W *et al* 2018 Landslide susceptibility modeling based on gis and novel bagging-based kernel logistic regression *Appl. Sci.* **8** 2540
[27] Truong X L, Mitamura M, Kono Y, Raghavan V, Yonezawa G, Truong X Q, Do T H, Bui D T and Lee S 2018 Enhancing prediction performance of landslide susceptibility model using hybrid machine learning approach of bagging ensemble and logistic model tree *Appl. Sci.* **8** 1046
[28] Bronstein M M, Bruna J, Cohen T, and Veličković P 2021 Geometric deep learning: grids, groups, graphs, geodesics, and gauges (arXiv:2104.13478)
[29] Kipf T N and Welling M 2017 Semi-supervised classification with graph convolutional networks *5th Int. Conf. on Learning Representations, ICLR 2017*, (*Toulon, France, April 24-26, 2017*), (*Conf. Track Proc.*) (available at: OpenReview.net)
[30] Veličković P, Cucurull G, Casanova A, Romero A, Lio P and Bengio Y 2017 Graph attention networks (arXiv:1710.10903)
[31] Fan C, Zhang C, Yahja A and Mostafavi A 2021 Disaster city digital twin: a vision for integrating artificial and human intelligence for disaster management *Int. J. Inf. Manage.* **56** 102049
[32] Wang S, Li Y, Zhang J, Meng Q, Meng L and Gao F 2020 Pm2. 5-gnn: a domain knowledge enhanced graph neural network for pm2. 5 forecasting *Proc. 28th Int. Conf. on Advances in Geographic Information Systems* pp 163–6
[33] Dees B S, Xu Y L, Constantinides A G and Mandic D P 2021 Graph theory for metro traffic modelling *2021 Int. Joint Conf. on Neural Networks (IJCNN)* (IEEE) pp 1–5
[34] Farahmand H, Xu Y and Mostafavi A 2023 A spatial–temporal graph deep learning model for urban flood nowcasting leveraging heterogeneous community features *Sci. Rep.* **13** 6768
[35] McBrearty I W and Beroza G C 2022 Earthquake location and magnitude estimation with graph neural networks *2022 IEEE Int. Conf. on Image Processing (ICIP)* (IEEE) pp 3858–62
[36] Norwegian Water Resources and Energy Directorate Skredhendelser - WMS 2013 (available at: https://nve.geodataonline.no/arcgis/services/SkredHendelser/MapServer/WMSServer?request = GetCapabilities&service = WMS) This work is licensed under the Norwegian Licence for Open Government Data (NLOD). To view a copy of this license, visit https://data.norge.no/nlod/en/2.0 (Accessed 1 May 2023)
[37] Norwegian Water Resources and Energy Directorate 2010 Avalanche steepness map - WMS (available at: https://nve.geodataonline.no/arcgis/services/Bratthet/MapServer/WMSServer?request = GetCapabilities&service = WMS) This work is licensed under the Norwegian Licence for Open Government Data (NLOD). To view a copy of this license, visit https://data.norge.no/nlod/en/2.0 (Accessed 1 May 2023)
[38] Wilde M, Günther A, Reichenbach P, Malet J-P and Hervás J 2018 Pan-european landslide susceptibility mapping: elsus version 2 *J. Maps* **14** 97–104
[39] Günther A, Van Den Eeckhaut M, Malet J-P, Reichenbach P and Hervás J 2014 Climate-physiographically differentiated pan-european landslide susceptibility assessment using spatial multi-criteria evaluation and transnational landslide information *Geomorphology* **224** 69–85
[40] Duscher K, Günther A, Richts A, Clos P, Philipp U and Struckmeier W 2015 The gis layers of the "international hydrogeological map of Europe 1: 1,500,000" in a vector format *Hydrogeol. J.* **23** 1867
[41] Arino O, Ramos Perez J J, Kalogirou V, Bontemps S, Defourny P and Van Bogaert E (European Space Agency (ESA) & Université catholique de Louvain (UCL)) 2012 Global Land Cover Map for 2009 (GlobCover 2009) [dataset] *PANGAEA* (https://doi.org/10.1594/PANGAEA.787668)
[42] Reuter H 2007 A Europe-wide digital elevation model based on SRTM and Russian topographic contours *Data set and documentation for the contract 2007-4500049350* (BGR)

[43] Lussana C, Tveito O E, Dobler A and Tunheim K 2019 seNorge_2018, daily precipitation and temperature datasets over Norway *Earth Syst. Sci. Data* **11** 1531–51
[44] Saloranta T M 2012 Simulating snow maps for Norway: description and statistical evaluation of the senorge snow model *Cryosphere* **6** 1323–37
[45] Panagos P *et al* 2022 European soil data centre 2.0: soil data and knowledge in support of the eu policies *Eur. J. Soil Sci.* **73** e13315
[46] Panagos P, Van Liedekerke M, Jones A and Montanarella L 2012 European soil data centre: response to European policy support and public data requirements *Land Use Policy* **29** 329–38
[47] ESDAC 2023 European Soil Data Centre esdac.jrc.ec.europa.eu This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit (available at: https://creativecommons.org/licenses/by/4.0/) (Accessed 8 May 2023)
[48] Norwegian Meteorological Institute senorge (available at: https://thredds.met.no/thredds/catalog/senorge/catalog.html) This work is licensed under the Norwegian Licence for Open Government Data (NLOD). To view a copy of this license, visit https://data.norge.no/nlod/en/2.0 (Accessed 1 May 2023)
[49] Statistics Norway 2022 Urban settlements (available at: https://kartkatalog.geonorge.no/metadata/tettsteder-2022/28906bc8-a644-43d5-bbdb-33a217f458ef This work is licensed under the Norwegian Licence for Open Government Data (NLOD). To view a copy of this license, visit https://data.norge.no/nlod/no/1.0 (Accessed 1 May 2023)
[50] Norwegian Mapping Authority 2018 Inspire transport network road (available at: https://kartkatalog.geonorge.no/metadata/inspire-transport-network-road/a76704df- fdd4-4e09-92ac-a7ad1a77b328) This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/ (Accessed 1 May 2023)
[51] Hosmer D W, Lemeshow S and Sturdivant R X 2013 *Applied Logistic Regression* vol 398 (Wiley)
[52] Matoušek J *et al* 2003 *Using the Borsuk-Ulam Theorem: Lectures on Topological Methods in Combinatorics and Geometry* vol 2003 (Springer)
[53] Zhang M, Kalies W D, Scott Kelso J A and Tognoli E 2020 Topological portraits of multiscale coordination dynamics *J. Neurosci. Methods* **339** 108672
[54] Von Luxburg U 2007 A tutorial on spectral clustering *Stat. Comput.* **17** 395–416
[55] Chung F R K 1997 *Spectral Graph Theory* vol 92 (American Mathematical Society)
[56] Norwegian Water Resources and Energy Directorate 2020 Avalanche warnings (available at: www.varsom.no/en/avalanches/avalanche-warnings/) This work is licensed under the Norwegian Licence for Open Government Data (NLOD). To view a copy of this license, visit https://data.norge.no/nlod/en/2.0 (Accessed 20 June 2023)
[57] Norwegian Water Resources and Energy Directorate 2020 Flood and landslide warning service (available at: www.varsom.no/en/flood-and-landslide-warning-service/) This work is licensed under the Norwegian Licence for Open Government Data (NLOD). To view a copy of this license, visit https://data.norge.no/nlod/en/2.0 (Accessed 20 June 2023)
[58] Dimasaka J, Selvakumaran S and Marinoni A 2023 Near-real-time Country-wide Estimation of Susceptibility and Settlement Exposure from Norwegian mass movements via inter-graph representation learning *Zenodo* (https://doi.org/10.5281/zenodo.8099812)