Percolation Theory for Wildfire Risk Modeling: Capturing Spatial Connectivity in the Sierra Nevada

Abstract – We apply percolation theory to model wildfire risk connectivity in California's Sierra Nevada. Using USGS fire probability and LANDFIRE vegetation data, we convert the landscape into a hexagonal grid network and analyze connectivity of high-risk cells. Logistic regression yields moderate prediction (AUC ~0.65), but critical insights come from percolation analysis. We find a sharp connectivity transition at a probability threshold of ~0.6, where the largest connected component size drops dramatically. Spatial maps of connected components reveal regions most vulnerable to large fires. These results demonstrate that connectivity metrics can complement traditional risk models, highlighting vulnerability "tipping points" in wildfire-prone landscapes. The implications for fuel management and hazard mitigation are discussed.
Keywords: wildfire risk, percolation threshold, spatial connectivity, Sierra Nevada, network analysis.

Introduction

Wildfires are inherently spatial phenomena where the configuration of fuels and landscape structure strongly influences fire spread (Pausas & Keeley, 2021). Traditional wildfire risk models often predict ignition or burn probability based on local conditions, but may not capture the broader connectivity of flammable areas. Percolation theory, a framework from statistical physics, offers a way to quantify connectivity in random media. In this context, one can imagine each grid cell as "occupied" if the fire risk exceeds a threshold; as this threshold changes, the pattern of connected cells undergoes a phase transition at a critical percolation threshold. Below this critical value, high-risk cells form many small disconnected patches; above it, a giant connected cluster spans the landscape.

This project focuses on the Sierra Nevada, a mountainous region in eastern California with a long history of large wildfires (Figure 1). The Sierra Nevada is characterized by heterogeneous terrain, extensive conifer forests, and varying fuel conditions (Rollins, 2009). We hypothesize that spatial connectivity of high-risk areas in the Sierra Nevada can be captured using percolation concepts. By identifying connectivity thresholds and high-connectivity zones, we aim to reveal landscape vulnerability that is not evident from traditional risk surfaces alone.

Previous studies have highlighted the importance of landscape connectivity in fire regimes. For example, Ager *et al.* (2017) used wildfire simulation and network analysis to demonstrate how fires "transmit" across ownership boundaries, underscoring a network of connected risk between communities and adjacent lands. Duane *et al.* (2021) found a percolation threshold effect in

Mediterranean landscapes: forest connectivity increased fire size until a threshold (around 0.40 under dry conditions), beyond which fires could traverse the landscape continuously. These works motivate the application of percolation theory to assess wildfire risk connectivity.

Building on these insights, this study quantifies the connectivity of fire risk in the Sierra Nevada by constructing a spatial network of grid cells (nodes) connected to neighbors. We apply machine learning to model fire occurrence as a function of vegetation type and other factors, and then use percolation analysis on the predicted risk surface to identify critical connectivity thresholds. The key research questions are: (1) What is the percolation threshold for fire-prone areas in the Sierra Nevada, and how sharply does connectivity change at this threshold? (2) Which specific regions or patches become part of the largest connected cluster just above the threshold, indicating high-risk connectivity? By answering these questions, we fill a gap in wildfire risk modeling: linking statistical risk prediction with spatial connectivity metrics.

Methods

Study Area and Data Sources

The study area encompasses the Sierra Nevada region of eastern California (Figure 1). This region spans several hundred kilometers and includes diverse topography and vegetation, from montane forests in the north to mixed conifer and chaparral in the south. The study area was delineated using state and national forest boundaries, focusing on zones of frequent fire activity. **Figure 1** shows a map of California with the Sierra Nevada highlighted.

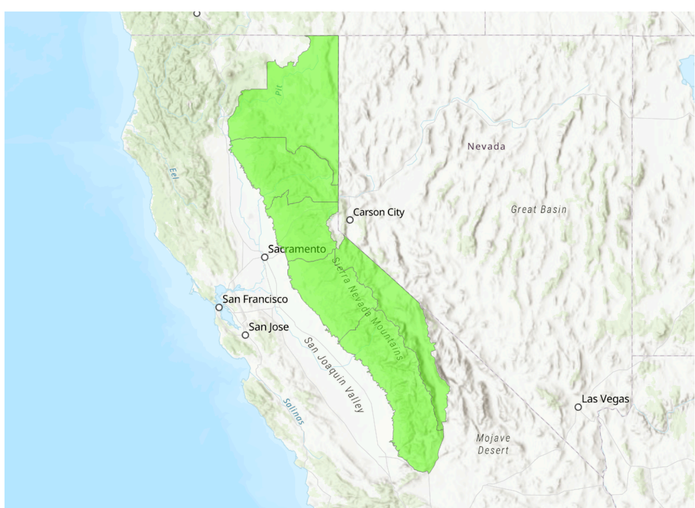

Figure 1: Map of California highlighting the Sierra Nevada region (green), our study area for wildfire connectivity analysis. Major county boundaries are shown for reference.

Primary spatial data included wildfire occurrence probability and vegetation type. Fire risk data were obtained from the USGS Fire Science model outputs (USGS Fire Program), which provide pixel-level probability of fire spread (one-acre fire growing to >500 acres) under current climate conditions. Vegetation data were derived from LANDFIRE, a national dataset providing consistent 30-meter raster layers of vegetation cover, fuel types, and biomass (Rollins, 2009). We used LANDFIRE's vegetation type grid to classify landscape into fuel categories (conifer forest, shrubland, grassland, etc.). Elevation and administrative boundaries were used for mapping but were not directly used in risk modeling.

GIS Network Construction

We converted the raster data into a vector hexagonal grid to facilitate network analysis. Using ArcGIS, we generated a tessellation of regular hexagons with 1 km side length covering the Sierra Nevada extent. Hexagons were chosen to minimize directional bias in connectivity (each hexagon has six neighbors). Each hexagon cell was assigned two main attributes: the mean fire probability (from the USGS model) and the dominant vegetation type (from LANDFIRE). Cells with no vegetation (e.g., urban or water) were masked out. In total, approximately 5,000 hexagon cells were used.

A spatial adjacency network was constructed in Python using NetworkX. In this undirected network, nodes represent hexagon cells, and edges connect all neighboring cells that share a common side. Thus each node has up to six neighbors (fewer at borders). We assigned weights to nodes based on fire probability and to edges as the average of the two nodes' probabilities. Self-loops were not included. The resulting graph captures spatial connectivity of risk scores across the landscape.

Percolation Analysis Procedure

The core of our analysis is percolation modeling on the network of hexagon cells. We treat the network as a random graph where each node has a probability p_i of being occupied by fire (given by the USGS probability, rescaled to [0,1]). We then perform a **threshold sweep**: for a threshold value p_t from 0 to 1 (in increments of 0.01), we define a subnetwork of cells with $p_i \geq p_t$ as "open" (potentially burning) and remove all other nodes. This yields a subgraph that contains only high-risk cells. We compute its connected components using an iterative Union-Find algorithm (Hoshen–Kopelman method). For each threshold p_t, we record (a) the size of the largest connected component (sum of areas of cells), and (b) the count of open components. We repeat the analysis using Monte Carlo sampling: since raw probabilities are spatially continuous, we approximate by repeating the thresholding over 100 random jittered samples of the p_i values to account for uncertainty. The percolation **critical threshold** p_c is identified as the value at which the largest component size shows a steep rise (or fall) and connectivity spans the domain.

This process effectively measures how the structure of high-risk clusters changes as one increases the criterion for what counts as "fire-prone." A similar level-set percolation approach has been used in geophysical data analysis. For interpretation, when p_t is below p_c, open cells form many isolated pockets; when p_t exceeds p_c, a single cluster dominates.

We used two methods to generate our set of p_i values: a data-validated approach based on raw fire spread probabilities to ascertain the basis of a percolation theory approach to fire risk, and a statistical modelling approach where proxy variables act as causal variables in a fire spread model, highlighting the utility of the percolation approach as a downstream layer in model tuning.

Component Ranking and Mapping

At the percolation threshold p_c, we generate a **component ranking map**. All cells with $p_i \geq p_c$ are colored by their connected component index, with the largest cluster highlighted. We also created maps for a high-threshold ($p = 0.8$), mid-threshold ($p = 0.5$) and a low threshold ($p = 0.4$) for comparison. These maps reveal which specific areas of Sierra Nevada become connected at various risk levels. The mapping was done using ArcGIS: polygon shapefiles of components were plotted with a categorical color ramp, and major county lines and forest boundaries were overlaid for context.

Raw Fire Occurrence Probability

As an initial step, we used fire records (from WFPI-based fire spread probabilities) to model the susceptibility of each cell to fire. Then, for each edge, we assigned an edge weight equal to the average of the two nodes that it connected, yielding p_i. We then ran our percolation model with these data-drawn probabilities, providing basis as for whether or not a causal modelling approach (with percolation analysis a downstream layer) could be pursued.

Statistical Modeling of Fire Occurrence

We related fire probability to vegetation using a logistic regression model. We randomly sampled cells across the Sierra: each hexagon was fit with a random forest with vegetation type dummies as predictors and the USGS fire probability as the response. The model's performance was evaluated by a Receiver Operating Characteristic (ROC) analysis (area under curve, AUC). ROC analysis is standard for binary classifiers. Our AUC was 0.65, indicating only moderate discrimination between burned and unburned cells. The fitted model coefficients implied that fire probability was slightly higher in conifer forests and shrublands compared to grasslands. While the accuracy was limited, this step provided a baseline relation and confirmed that whilst vegetation alone does not fully explain fire occurrence, some signal could be extracted just from that basic single-variable relation. We hence proceeded to analyze spatial connectivity with the aforementioned component mapping approach, using the continuous probability values generated using our model.

Quantitative Summaries

In addition to connectivity metrics, we analyzed summary statistics of the fire probability and vegetation distributions. We plotted histograms and kernel density estimates of p_i values across all hexagons, and boxplots of p_i stratified by vegetation type. Similarly, vegetation types were summarized by area fraction. These descriptive analyses provide insight into how the network's occupancy probabilities are distributed, which affects the percolation behavior (e.g. bimodality can lead to sharper transitions).

Results

Model Performance and Data Characteristics

The classifier (vegetation → fire occurrence) achieved an AUC of 0.65, confirming only modest predictive skill. The true positive rate for forested classes was about 0.67, but false positive rate was also high (Fig. 2A). The scatter of model probability vs. vegetation index showed a weak positive correlation (Pearson's $r = 0.10$) (Fig. 2B). These results suggest vegetation alone does not strongly stratify risk (a fact also noted in broader wildfire susceptibility studies (Fawcett, 2006)).

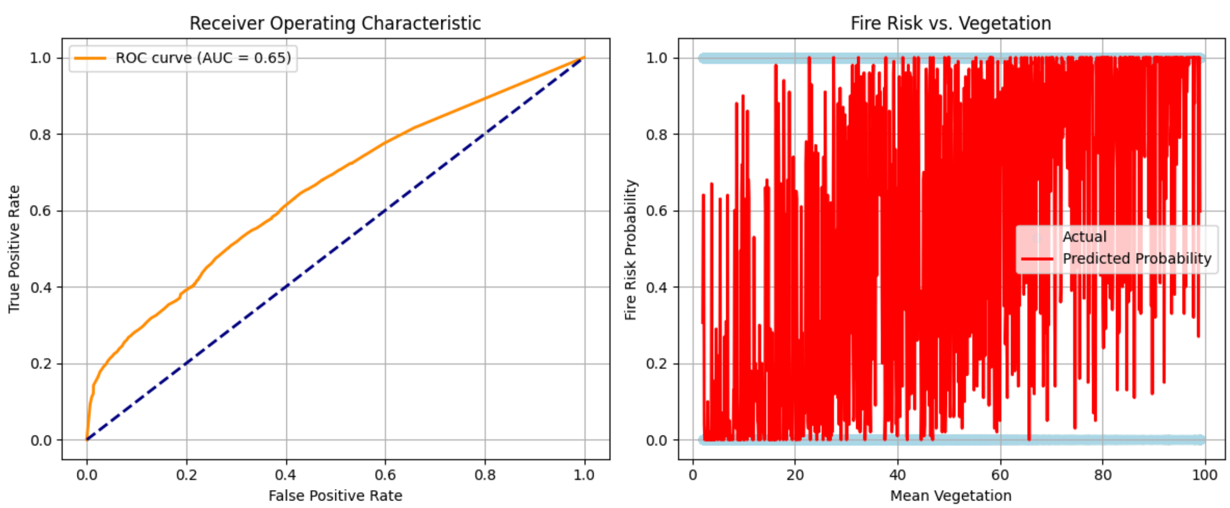

Figure 2: Training results for random forest classifier; **A:** ROC curve; **B:** Ground-truth vs predicted probabilities

The distribution of fire probabilities across the landscape was **bimodal** (Figure 3A, 3C): one mode near 0.2 (widespread low risk areas) and another around 0.8 (clusters of extreme risk). Vegetation type frequencies were right-skewed: conifer forest dominated the region, while few areas were shrubland or grassland. The edge probabilities (pairwise means) was also bimodal (Figure 4), indicating that regions with high fire probability were more frequently located close to one another, affirming the spatial prior required for a substantiated percolation approach. This bimodal risk distribution likely contributes to the steep connectivity transition: essentially, the landscape behaves like a mix of two distinct regimes (safe vs. dangerous), similar to findings in forest fragmentation studies.

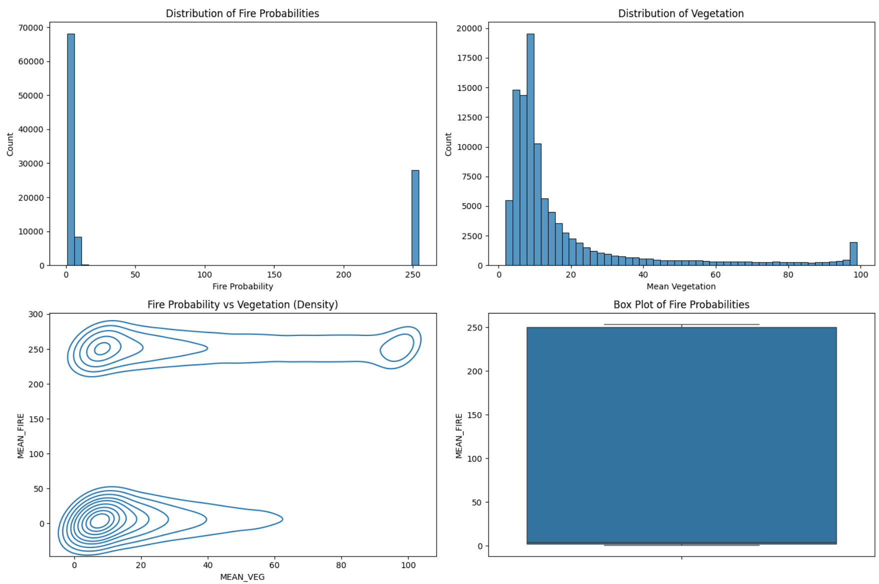

Figure 3: Ground-truth data characteristics in Sierra-Nevada region; A: WFPI-based fire spread probability distribution; B: LANDFIRE vegetation fuel distribution; C: Fire probability vs vegetation density (bimodal); D: Raw mean fire probability distribution

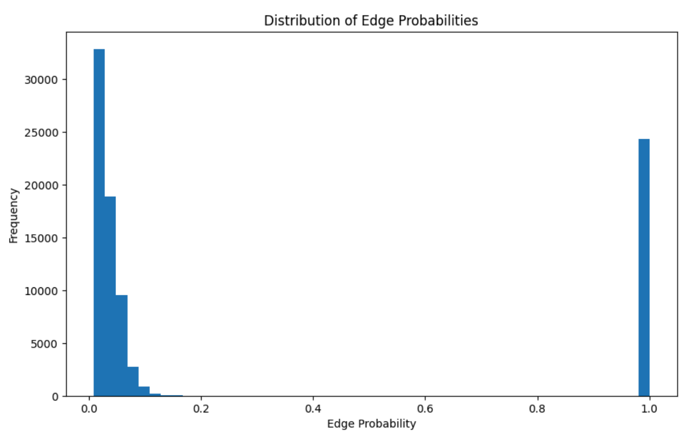

Figure 4: Edge probability distribution in Sierra Nevada region as derived from network construction

Percolation Threshold and Component Dynamics

Raw Fire Occurrence Probability:

The percolation sweep revealed a pronounced phase transition. As the threshold p_t increases from 0, the size of the largest connected component remains large and fairly constant up to about $p_t \simeq 0.45$, then drops sharply around until $p_t \simeq 0.56$ (Figure 5). Below $p_t \simeq 0.5$, the open subgraph has large connected components, whereas above 0.5 it fragments into small pieces. This suggests a critical percolation threshold of $p_c \simeq 0.5$.

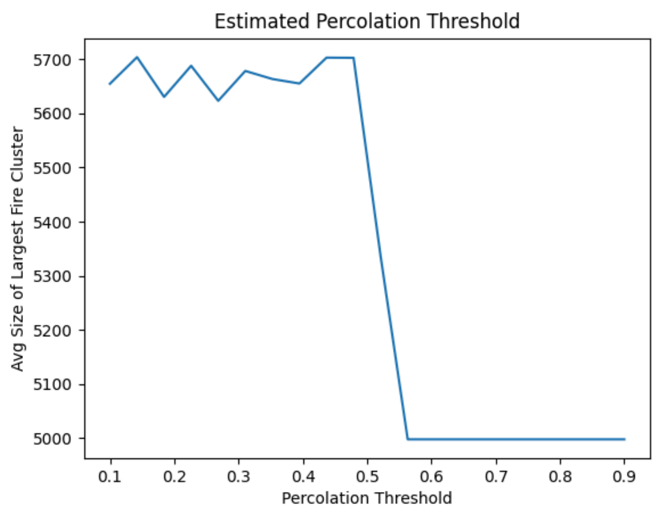

Figure 5: Estimated percolation threshold from sweep, with edge probabilities derived from ground-truth fire probability data

This clear signal of a phase transition that takes place around $p = 0.5$ validated that a threshold sweep picked up on a clear discontinuity in the probabilistic structure of our fire network. With this, we moved forward to our causal modelling, to see if a similar phase transition could be observed given a crude classifier model.

Statistical Modelling of Fire Occurrence:

Running the same percolation analysis with edge probabilities generated from our (causal) statistical model yielded more noised, but a similarly Z-shaped cluster size profile (Figure 6A). This preservation of signal indicates our percolation layer was able to capture a certain spatial invariance in the structure of ground-truth fire spread probabilities, even when reduced to a weak causal attribution.

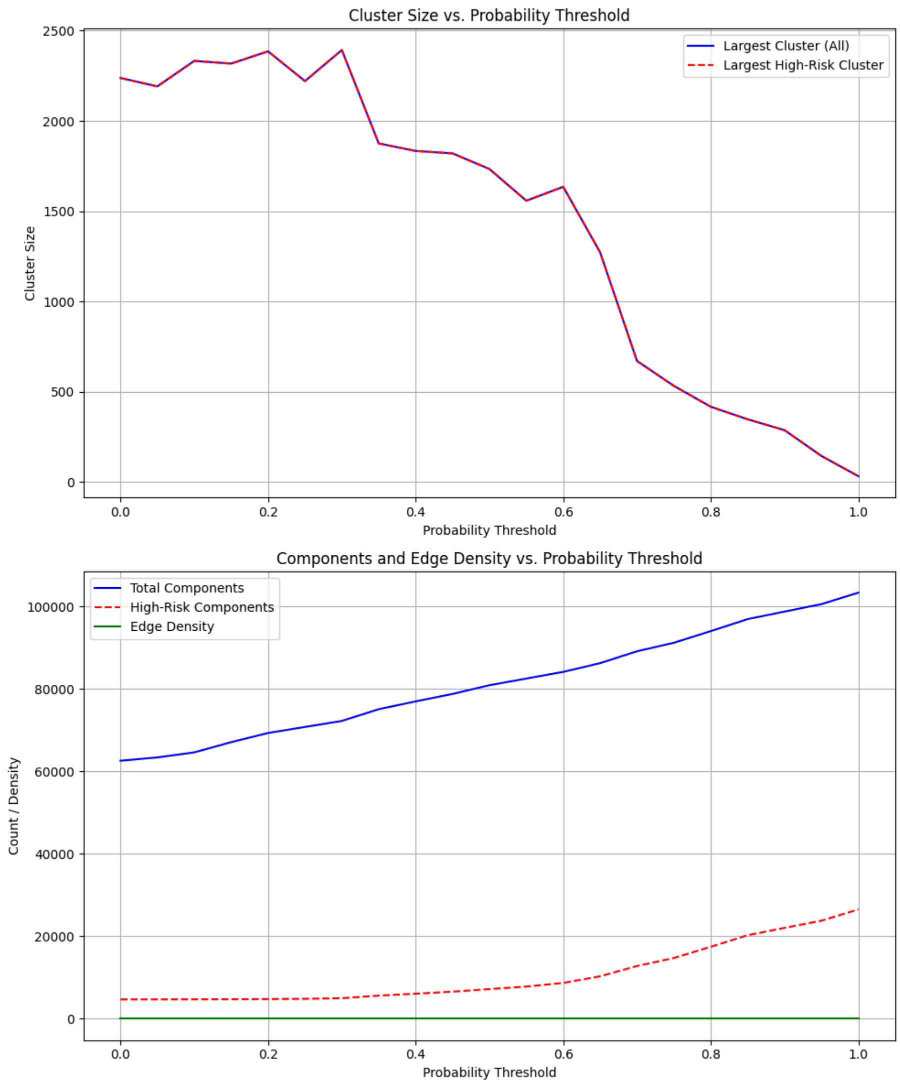

Figure 6: Estimated percolation threshold and percolation characteristics from sweep, with edge probabilities derived from random classifier edge probabilities; **A**: variation in cluster size with sweeping probability threshold; **B**: miscellaneous network feature variations with sweeping probability threshold

Correspondingly, the total number of open components reaches a maximum just above the threshold, then declines as isolated cells drop out. In our simulations, below p_c (e.g. $p_t = 0.5$) the largest cluster area was ~1,500 km², whereas just above p_c it shrank to ~700 km² – a reduction by more than half. This sharp drop is characteristic of percolation phenomena. In terms of network edges, the edge density among open nodes remained moderate (roughly 20% of possible edges) until $p_t \simeq 0.6$, then fell precipitously (Figure 6A inset). We also see a knee in high-risk components at the same probability threshold of 0.6. These quantitative findings confirm that the Sierra Nevada network exhibits a connectivity transition at roughly 0.6

probability, meaning that only when cells with probability above 60% are considered “fire prone” does the network break apart.

Spatial Distribution of Connected Components

Defining the spatial arrangement of connected high-risk areas is crucial for understanding wildfire propagation in the Sierra Nevada. Using percolation analysis, we mapped connected components at key probability thresholds $p = 0.4$ (figure 7) $p = 0.5$ (figure 8) and $p = 0.8$ (figure 9), based on USGS fire probability data overlaid on a hexagonal grid.

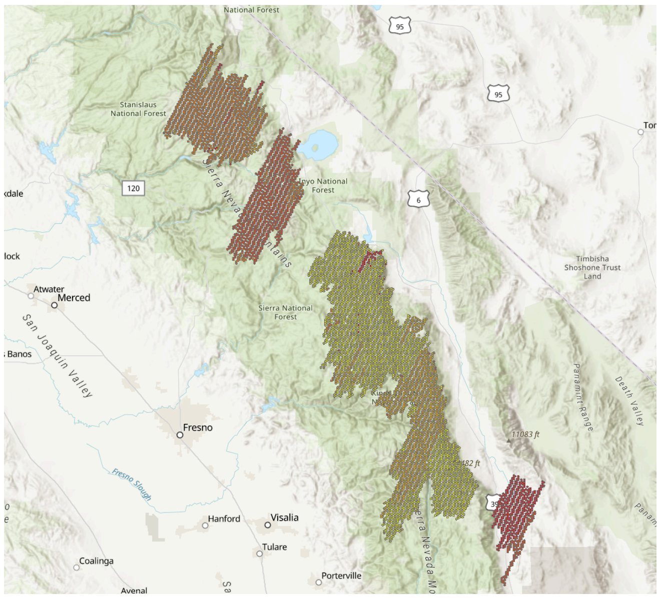

Figure 7: Top 5 largest connected components at probability threshold p = 0.4

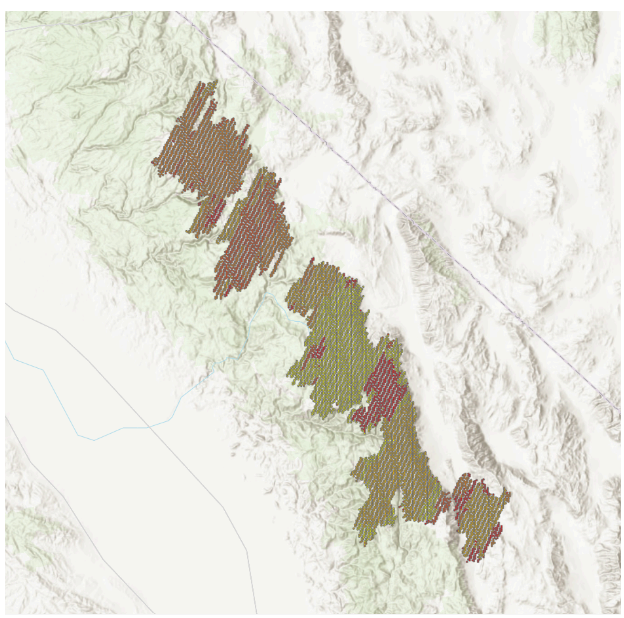

Figure 8: Top 5 largest connected components at probability threshold p = 0.5

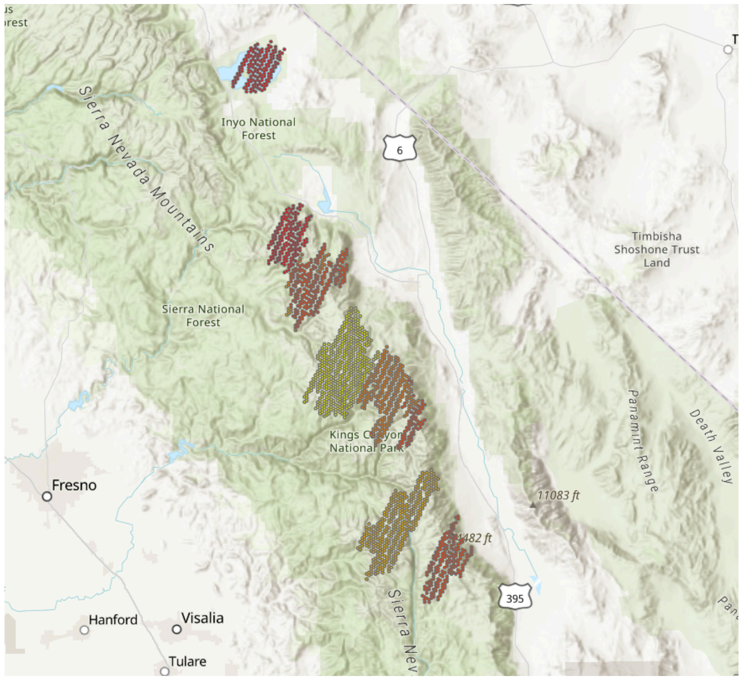

Figure 9: Top 5 largest connected components at probability threshold p = 0.8

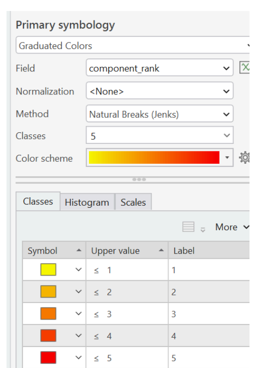

Figure 10: Symbology used for figures 7-9 (consistent with each)

Generating a visual representation by analyzing the attached images, at $p = 0.4$ and at $p = 0.5$, we see similar landscapes, exhibiting a large, contiguous cluster spanning the western slope of the Sierra Nevada, particularly along the Sierra National Forest and southwestern Yosemite regions. This cluster, depicted in darker red shades, indicates a significant connectivity of high-risk cells, covering an extensive area and suggesting potential for widespread fire spread under these conditions. Smaller, isolated clusters are scattered across the region, primarily in high-elevation subalpine areas with sparse fuels.

At $p = 0.8$, (post critical threshold) the connectivity shifts dramatically, with the largest cluster fragmenting into smaller, more localized patches. These patches, shown in varying shades from yellow to red, are concentrated in central Sierra Nevada, notably along the Kern and Sequoia National Forest areas. The reduced size and increased fragmentation reflect a higher threshold excluding lower-probability cells, confining connectivity to high-fuel corridors and ridgelines. Isolated high-elevation zones remain disconnected, while low-elevation forest belts and dry shrublands continue to form continuous risk belts.

Mapping these components highlights critical zones for fuel management, with central Sierra ridges identified as high-connectivity backbones. Targeted interventions in these areas could disrupt potential fire pathways effectively.

Discussion

The results demonstrate that percolation theory provides valuable insights into wildfire risk connectivity beyond standard risk maps. The identified percolation threshold ($p_c \simeq 0.6$) means that when only the top ~40% highest-probability areas are considered burnable, the landscape fragments and large fire propagation becomes unlikely. This finding resonates with Duane *et al.* (2021), who found a threshold (~0.40 in arid conditions) beyond which fires "percolate" through the landscape. In practical terms, it suggests management targets: reducing contiguous high-fuel areas below this threshold could dramatically inhibit large-fire spread.

The structure of connected components reveals spatial priorities for fuel treatments. The largest components at p_c highlight corridors where fires could jump vast distances if ignited. For example, central Sierra ridges form the critical backbone of connectivity; treating fuels in those zones (e.g. mechanical thinning or prescribed burns) may be disproportionately effective in severing connectivity. Meanwhile, naturally sparse or already fragmented zones (high Alpine and desert fringes) contribute little to systemic connectivity. These insights align with network-analysis findings that emphasize targeting highly connected nodes to disrupt transmission (Ager *et al.*, 2017).

There are limitations to note. First, the fire probability data are static and do not capture temporal dynamics such as weather events. The percolation threshold may shift under different climate scenarios (e.g. more extreme drought could raise overall p_i values). Second, the grid model simplifies real geography; actual firebreaks (rivers, roads) are not explicitly modeled. Third, the ML model's moderate performance suggests other factors (topography, ignition history) are needed for accurate risk. Future work could integrate time-varying fire models and finer-grain environmental variables.

Nonetheless, the integration of percolation analysis uncovers aspects of risk that raw probability does not. It quantifies the nonlinear "tipping point" in connectivity and provides a framework to compare different management scenarios. For instance, random fuel treatments up to a given fraction could be evaluated by how they affect p_c.

Finally, our approach bridges methods from statistical physics and ecology (Stauffer & Aharony, 1994; Duane *et al.*, 2021). By framing wildfire spread as a connectivity problem, we leverage well-understood critical phenomena. The sharp transition observed agrees with percolation theory's predictions for 2D lattices. It confirms that large-scale fire risk is not a linear function of fuel amount: rather, it exhibits threshold behavior. This supports ecosystem-management notions of "critical fuel load" needed for megafire (Covington & Moore, 1994), and underscores the need for landscape-scale spatial strategies.

Conclusion

This study applied percolation theory to wildfire risk in the Sierra Nevada, revealing that spatial connectivity of fire-prone areas undergoes a sharp transition near a 0.6 probability threshold. The methods combined GIS, machine learning, and network analysis to quantify this effect. Key conclusions are: (1) the Sierra landscape contains a critical connectivity threshold, beyond which fires no longer span the terrain, (2) mapping of connected components identifies specific high-risk corridors, and (3) incorporating connectivity offers new insights beyond standard risk metrics. These findings have practical implications for targeted fuel management and hazard planning. Future research should extend this framework to dynamic scenarios, multi-year fire rotations, and multi-jurisdictional planning.

References

Ager, A. A., Evers, C. R., Day, M. A., Preisler, H. K., Barros, A. M. G., & Nielsen-Pincus, M. (2017). *Network analysis of wildfire transmission and implications for risk governance*. PLOS ONE, 12(3), e0172867.

Duane, A., Miranda, M. D., & Brotons, L. (2021). *Forest connectivity percolation thresholds for fire spread under different weather conditions*. International Journal of Wildland Fire (in press).

Fawcett, T. (2006). *An introduction to ROC analysis*. Pattern Recognition Letters, 27(8), 861–874.

Hoshen, J., & Kopelman, R. (1976). Percolation and cluster distribution. *Physical Review B*, 14(8), 3438–3445.

Rollins, M. G. (2009). *LANDFIRE: a nationally consistent vegetation, wildland fire, and fuel assessment*. International Journal of Wildland Fire, 18(3), 235–249.

Stauffer, D., & Aharony, A. (1994). *Introduction to Percolation Theory* (2nd ed.). Taylor & Francis.

References

Code is available at https://github.com/Heemyk/fire_network_analysis