



Universiteit
Leiden

Master Computer Science

Assessing Complex Fire Spread Dynamics Using
Spatiotemporal, Network-Based and Contagion-
Based Models

Name: Mengjie Hu
Student ID: s3862690
Date: 11/11/2024
Specialisation: Data Science
1st supervisor: Sebastian Fajardo Bernal
2nd supervisor: Alberto Ceria
3rd supervisor: Frank Takes

Master's Thesis in Computer Science

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Abstract

Fire risk assessment remains challenging, as many existing models are data-intensive, computationally costly, or function as 'black boxes'. This thesis presents a framework that integrates a spatiotemporal network (Chronnet) with a Susceptible-Infected-Susceptible (SIS) epidemic model to reproduce the complex spatiotemporal dynamics of fire spread. Using Colombia as a case study with 12 years of VIIRS satellite data, the framework first transforms thermal anomaly detections into a directed, weighted network of sequential ignition events. The SIS model is then applied to compute steady-state burning probabilities, which we use as an indicator of fire risk. In the evaluation framework, a key contribution is the distinction between spontaneous ignitions and network-driven spread, ensuring evaluation focuses on dynamics the model is designed to capture. The results demonstrate that the Chronnet-SIS framework effectively captures these spatiotemporal patterns, consistently outperforming a PageRank centrality baseline in identifying high-risk areas, as measured by Average Recall (AR) and Normalized Discounted Cumulative Gain (nDCG). An ablation study confirms the model's robustness, showing that its explanatory power stems from modeling inter-cell contagion rather than just fire persistence. This research validates epidemic modeling as an interpretable tool for understanding complex fire dynamics, offering a valuable data-driven method for risk assessment.

Contents

1	Introduction	1
2	Related Work	3
3	Dataset	6
3.1	Data Sources	6
3.2	Preliminary Data Exploration	7
3.3	Spatial and Temporal Preprocessing	8
3.3.1	Hexagonal Grid Generation	8
3.3.2	Time Aggregation	10
4	Method and Experimental Design	10
4.1	Chronnet Construction	11
4.2	Network Based SIS Model for Fire Spread	12
4.2.1	Overview of the SIS Framework	14
4.2.2	Steady-State Solution of the SIS Model	15
4.2.3	Fire Spread Threshold	17
4.3	Experiment Design	18
4.3.1	Experimental Setup and Network Characterization	18
4.3.2	Modeling and Evaluation Framework	21
5	Results and Analysis	25
5.1	Method Validation	26
5.1.1	Chronnet Resolution Effects on fire Spread Patterns	26
5.1.2	Distinguishing Spontaneous and Network-Driven Fire Patterns	28
5.1.3	Robustness of the Ground Truth Definition	28
5.2	Chronnet Network Characterization	30
5.2.1	Degree and Strength Distributions of Chronnet	30
5.2.2	Topological Metrics of Top 2 SCCs	32
5.2.3	Relation between node metrics and fire behavior	33
5.3	SIS Model Performance Analysis	34
5.3.1	Experimental Results Across Different τ	35
5.3.2	Ranking Performance Evaluation using nDCG	38
5.3.3	Spatial Distribution of Fire Risk	40
5.3.4	Impact of Self-Loops on SIS Model Performance	42
6	Conclusion	44
6.1	Additional Figures	47
6.2	Code Availability	47

1 Introduction

Fires are a growing global concern, with increasing frequency and intensity driven by climate change and human activity. Over long periods, fire has been a fundamental ecological and evolutionary process, regulating vegetation regeneration, nutrient cycling, biodiversity patterns and other ecosystem functions [1]. However, the ability of fires to spread across wide regions remains a serious risk. These fires cause immense environmental destruction, contributing to carbon emissions, soil degradation, and biodiversity loss. Hence understanding and assessing the risk of the complex spatiotemporal spread of these fires is important for developing effective prevention and mitigation strategies.

Existing fire spread models, such as physics-based [2], semi-empirical models [3] and simulation models [4]. These method often demand precise environmental data, such as wind speed and detailed geographic and vegetation data (i.e, fuel type), which makes them making them challenging to apply in practice. More advanced graph-based deep learning models [5] [6] can be computationally intensive and difficult to interpret due to their ‘black-box’ nature. A promising alternative is to investigate whether a simple network-based spreading process, such as the Susceptible-Infected-Susceptible (SIS) model from epidemiology, can reproduce the complex spatiotemporal dynamics of fire spread. Although previous research [7] [8] has explored the theoretical use of epidemic-based models for fire dynamics, their application to real-world, recurrent fire data for risk assessment has been limited. This research addresses this critical gap by developing a data-driven modeling approach that can extract dynamic spread patterns from historical observations.

Specifically, we combine two methodologies to estimate fire risk. We first employ Chronnet [9], a method that constructs a spatiotemporal network from historical data to model how fires spread between locations over time. It involves linking spatial grid cells in the chronological order of their occurrences. This network then provides the structure for a Susceptible-Infected-Susceptible (SIS) model [10], an epidemic framework where each grid cell (or node) transitions between a burning (infected) state and a non-burning (susceptible) state. Using this combined approach, we estimate fire risk, which is defined as the probability of a grid cell being burning state. This allow us to identify critical regions for intervention.

To demonstrate and validate this framework, we apply it to a case study in Colombia, a nation where understanding fire spread is vital for its ecological health and safety. The country presents a challenge for fire management, its wide range of ecosystems and high biological diversity make nationwide fire monitoring difficult. This challenge is worsened by consistently high fire activity, with occurrences concentrated in the Orinoquia, Amazon, and Caribbean

regions [11]. In recent years, satellite-based thermal anomaly detection data from MODIS and VIIRS have become essential for locating ignition points and identifying high-density fire regions [12]. However, most methods using these datasets primarily focus on visualizing current fire activity rather than modeling how fires spread dynamically across space and time. This limits their utility for fire risk assessment and preventive management strategies. Our modeling approach can overcome this limitation by capturing the dynamics of fire propagation over time. To evaluate our model’s effectiveness, we benchmark its performance against PageRank [13], a well-known graph centrality algorithm that measures the importance of nodes within a network.

This thesis aims to address the following research questions:

1. To what extent can a Chronnet-based network, constructed from historical satellite data, capture the spatiotemporal dynamics of fire spread in Colombia?
2. How can a network-based Susceptible-Infected-Susceptible (SIS) model be effectively applied to this network to assess fire risk?
3. How does the SIS model compare to a baseline like PageRank in ranking high-risk areas?

This thesis makes the following key contributions. It presents a modeling framework that combines Chronnet and a network-based SIS model to analyze the spatiotemporal spread of fires, and demonstrates its application to real-world fire data from Colombia. A major contribution is the important distinction between spontaneous fires and network-driven ones, which is a necessary step for correctly evaluating any propagation-focused model. The reason for this separation is to align the evaluation with the model’s actual capabilities. Spontaneous fire represent new ignitions, while network-driven fires capture propagation within the system. In this framework, network-driven dynamics include both inter-cell spread (between neighboring cells) and intra-cell persistence (self-loops, continued burning within the same cell). By adopting this definition, model evaluation focuses on the propagation dynamics that Chronnet-SIS is designed to capture. We also show through ablation study that SIS model’s performance primarily stems from its ability to capture the inter-cell fire spread pattern, rather the intra-cell persistence (self-loops). Furthermore, the thesis also includes a systematic analysis of how different grid sizes affect network structure and model results, helping to find a suitable resolution for fire modeling. Finally, the proposed model shows a clear advantage over the PageRank baseline in identifying high-risk areas task. This result validates the approach of modeling fire as a dynamic contagion process.

This thesis is organized as follows. Section 2, *Related Work*, reviews existing fire spread models, with a focus on graph-based and epidemic-inspired approaches that inform our method. Section 3, *Dataset*, introduces the dataset and describes key preprocessing steps, including hexagonal grid generation and temporal aggregation. Section 4, *Method and Experimental Design*, presents the proposed modeling framework, which combines Chronnet with a network based SIS model, and explains the experimental design and evaluation metrics. Section 5, *Results and Analysis*, reports the experimental results and provides analysis. Section 6, *Conclusion*, concludes the thesis by summarizing the contributions, discussing the findings, and outlining possible directions for future work.

2 Related Work

In this part, we review how other works have modeled fire spread, with a particular focus on graph-based approaches. First we discuss methods for graph construction, and then turn to models of the spreading dynamics, focusing especially on epidemic based models. The papers included were selected because they either inform our own methodological choices or highlight gaps that our work seeks to address.

Fire spread are complex spatiotemporal phenomena influenced by weather, land features, vegetation, and human activities [14]. Conventional fire model include statistical models [15], physics-based models [2], semi-empirical models [3], simulation models [4] and machine learning model. Among these, graph-based model, ranging from traditional network simulations to modern graph-based machine learning, have emerged as a prominent approach for its effectiveness in capturing the spatial and temporal dynamics of fire spread.

Graph-based model often represent complex fire spread process in a network form. The graph construction method plays an important role in fire spread modeling. One major method is to represent the physical landscape as a network of nodes connected by edges that represent potential fire transmission pathway. One early but influential example is Finney’s study [16]. In his work, the terrain is divided into regularly spaced nodes (square grids) with edges weighted by the time it takes fire to travel between nodes. Fire spread simulation then can be treated to find minimum travel time paths on this graph from the ignition point. This method is computationally efficient and can capture directional effects like winds. Encinas’ work [17] additionally shows that using hexagonal grids could help overcome the limitations of using square grids as spatial representation elements. Later studies build to improve spatial graph representation by using irregular mesh or graph that better match

terrain and fuel heterogeneity. For example, Johnston [18] introduced an irregular grid to minimize the error introduced by the regular grid-based model. Stepanov and Smith [19] use the Delaunay triangulation to represent the landscape and then applying shortest-path algorithms to propagate fire across the mesh. Hajian [20] further employed Delaunay triangulation in their network construction and treated fire spread as a stochastic shortest-path problem on a graph. While these methods are good at integrating fire physics via edge weights by adjusting the spatial representation method, they suffer from the increasing complexity of constructing and validating of the graph. For example, Jiang [5] mention that triangulating the landscape and calibrating edge travel times require careful preprocessing. Alternatively, Ferreira [9] proposed Chronnet model which focus on extracting fire spread patterns directly from spatiotemporal fire detection data, without assumption about physical fuel or terrain properties. It offers a computationally efficient way to construct a directed weighted network where nodes represent spatial grid cells and edges represent the consecutive occurrences of fire events between nodes. This approach can capture actual patterns of fire propagation across space and time, providing a more robust framework for fire spread modeling. We build on this work to further validate the use of Chronnet in fire spread studies and assess its modeling capability.

Building on the graph representations introduced above, a key challenge lies in how to model the actual fire spread dynamics over these structures. Traditional physics-based models often rely on precise parameters (like wind speed, slope, and humidity) to simulate fire behavior, but they’re usually hard to apply in areas with limited data. Latest graph-based deep learning models [5] [6] might perform well in fire simulation, but their interoperability is low, and they require a lot of training data and computational resources (and may also lack physical transparency). Alternatively, fire spreading can be modeled using epidemic models, which are originally designed to simulate the transmission of diseases. The idea is that fire spread is like disease transmission, which follows a process of local interactions across a network. Given the similarity between fire spread and disease transmission, it is natural to consider using epidemic model to describe fire dynamics. Indeed, because epidemic models are well-suited for modeling such network-based spreading processes, they have been adopted in a wide range of domains beyond epidemiology, such as modeling information spread on social networks [21], contagion in financial systems [22], and traffic congestion in air transportation networks [23]. This also inspired the use of epidemic models to simulate fire spread. More importantly, epidemic models have advantages in simplicity, computational efficiency, and interoperability compared to others.

The work presented in [10] introduces the two most general epidemic mod-

els and their variants: the Susceptible-Infected-Susceptible (SIS) and Susceptible-Infected-Removed (SIR) models. It then further extends these models to networked scenarios using mean-field approximations. In the SIS model, individuals can transition from the infected state back to the susceptible state, allowing repeated infections over time. In contrast, the SIR model assumes that once infected individuals recover, they gain permanent immunity. These models serve as the foundation for analyzing contagion dynamics in various systems. Building on this epidemic framework, researchers have explored identifying influential nodes in the network-based epidemic system to help in hindering epidemics. Methods can be grouped into three types. The first group focuses on global structural features, such as K-shell index [24] and its variants [25]. The second uses local information like node degree or clustering coefficient [26] to estimate a node's influence. The third group considers the actual spreading process, including centrality metrics obtained from the steady state of some dynamics, such as random walk centrality [27]. In the context for our study, we follow the third approach, as it enables us to better capture the interplay between fire recurrence, spatial connectivity and dynamic contagion patterns.

While the identification of influential nodes has been well studied in general epidemic spreading processes, relatively few works have focused on applying epidemic-based frameworks to model and analyze real-world fire dynamics, and most of them remain theoretical. Generally, SIS-based frameworks are typically used to explore recurrent fire dynamics or to design control problems rather than to simulate a single fire event's spread. For example, Somers and Manchester [7] introduce a fire management approach inspired by the SIS (Susceptible-Infected-Susceptible) model. They treated a landscape as a graph, where each node represents a location that can transition between burning and non-burning states, like a SIS cycle. They develop an algorithm to generate "priority maps" for surveillance using positive system theory. In their work, the infection rate models the fire spread between connected cells, while the recovery rate refers to the rate at which burning areas become non-burning again. They use the SIS model's analytical properties to rank areas by how critical they are to overall fire spread. While their work shows that SIS model dynamics can effectively model fire spread for prioritization tasks, the approach is quite theoretical and is evaluated only on synthetic data. In contrast, SIR-based models are more suited for modeling fire spreading processes. In the SIR framework, once a location burns it transitions to a removed state that no longer burns, while it can be re-ignited in the SIS framework. Bosters et al. [8] introduce a stochastic compartmental model inspired by the SIR framework. In their approach, each grid cell transitions among three discrete states: unburned (susceptible), burning (infected), and burned-out (removed). Their model extends traditional lattice spread rules by using a wind-weighted and probabilistic neighborhood struc-

ture and can estimate the probability of each cell being in a given state at every time step. They demonstrate that SIR-based model can replicate known spread patterns. While both SIS and SIR models offer valuable insights into fire dynamics, existing applications remain mostly theoretical. There is a clear need for studies that bridge this gap by applying epidemic models to real-world fire data for risk assessment, particularly for those capable of modeling recurrent fire patterns. In our case, the SIS framework is more suitable than SIR because the spatial resolution we are working with is quite coarse and the temporal scale is 12 years, so a cell can experience multiple burning status within the study period.

3 Dataset

This section introduces the VIIRS 375m Active Fire dataset used for this study, covering thermal anomaly detection data in Colombia from 2012 to 2024, and describes the key preprocessing steps for network construction. These steps include splitting the study area with a hexagonal grid for spatial structure and aggregating fire events into 12-hour time windows. This temporal aggregation is crucial for addressing data sparsity and satellite observational biases.

3.1 Data Sources

We used the VIIRS 375m Active Fire product. This product provides global active fire detection with geographic location, date, confidence, additional information for each fire at 375m resolution. It comes from a satellite sensor called VIIRS, which is on the Suomi NPP and NOAA-20 satellites, managed by NASA and NOAA. It detects heat and fires on Earth using thermal imaging. This product works well with another fire detection tool called MODIS. Both tools are good at identifying hotspots, but VIIRS has better image resolution, meaning it can detect smaller fires more accurately and provides a clearer picture of large fires. Because of this, the 375m data is useful for fire management, such as real-time alerts, and for scientific research that requires detailed fire maps. It is one of the datasets widely used in fire spread research. The user guide can be found at this link: [VIIRS C2 AF-375m User Guide](#). The dataset can be downloaded here: [NASA FIRMS Download Page](#).

In our study, we only use the fire event data across Colombia from 2012 to 2024. We removed the low confidence fire events (keeping 94% of the fire events) to improve the accuracy of the dataset and to reduce potential false positives. The following are some key attributes we use in our study.

Table 1: Description of VIIRS 375m Active Fire Product Attributes

Attribute	Description
Latitude	Center of nominal 375 m fire pixel.
Longitude	Center of nominal 375 m fire pixel.
Acq. Time	Time of acquisition/overpass of the satellite (in UTC).
Confidence	An indicator of the quality of individual hotspot/fire pixels, assigned as low, nominal, or high. This value is based on several algorithm criteria including temperature anomalies, sun glint effects, and regional data quality. Low confidence pixels may be affected by sun glint (daytime) or South Atlantic Magnetic Anomaly (nighttime). High confidence pixels correspond to strong thermal anomalies.

3.2 Preliminary Data Exploration

Figure 1 and 2 show the monthly and quarterly fire count in Colombia respectively. We can observe that fires mostly occur in Quarter 1, revealing a strong seasonal trend in fire occurrences. Although our current model does not account for seasonality, the seasonal patterns suggest we could improve it in the future.

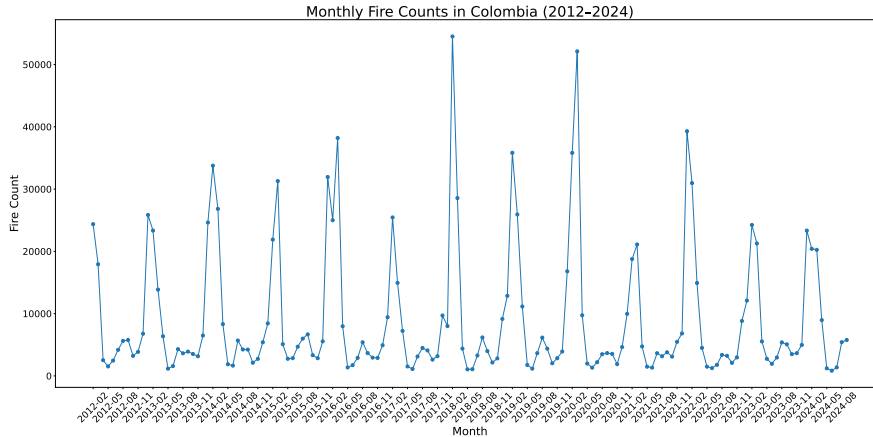


Figure 1: Monthly fire counts in Colombia (2012–2024).

Figure 3 shows the hourly fire counts. The peak hours are concentrated in two periods: morning (5–7 am) and evening (5–7 pm). This happens because the VIIRS satellite passes over Colombia (near the equator) twice a day. This phenomenon might suggest that the acquisition time (acq_time) is more

3 DATASET

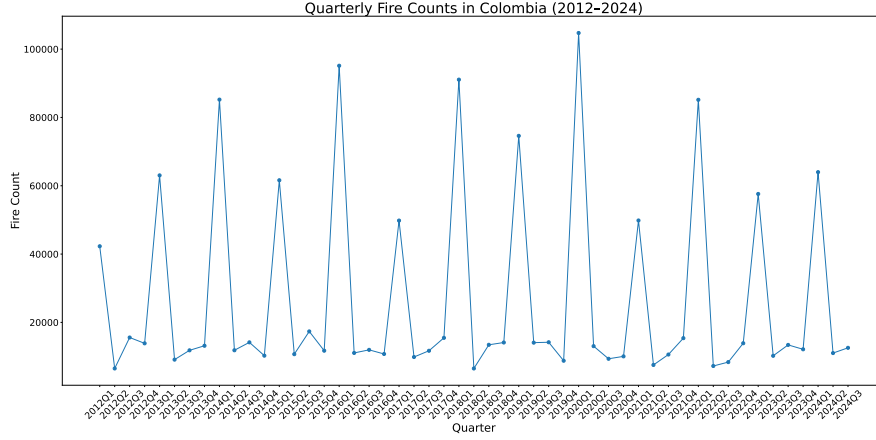


Figure 2: Quarterly Fire Counts in Colombia (2012–2024).

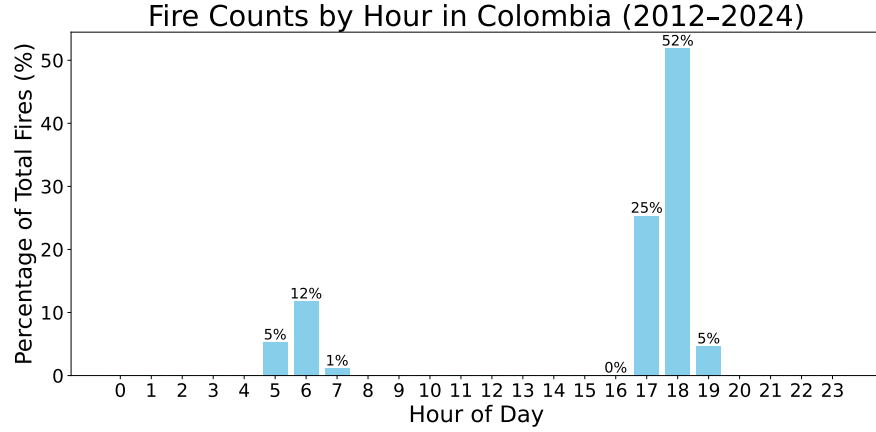


Figure 3: Hourly Fire Counts in Colombia (2012–2024).

influenced by the satellite observation schedule than by the actual temporal sequence of fire events.

3.3 Spatial and Temporal Preprocessing

3.3.1 Hexagonal Grid Generation

We use a hexagonal grid to divide the study area into spatial units. This helps build the Chronnet network. Hexagonal grids have advantages over traditional rectangular grids. They are more evenly spaced, providing more uniform coverage and reducing distortions. The side length of the hexagon is denoted by r , it determines the size of each hexagon, which helps to adjust the spatial resolution of ChronNet.

We create the hexagonal grid in five steps:

1. Boundary Extraction

We first get the bounding box of the study area from the GeoDataFrame. This box has the minimum and maximum x and y coordinates.

2. Grid Spacing

We use flat-topped hexagons. Each hexagon has a side length called `hex_size` r (in meters). The horizontal spacing is $1.5 \times \text{hex_size}$. The vertical spacing is $\sqrt{3} \times \text{hex_size}$. These values make sure the hexagons fit together neatly.

3. Center Point Generation

We create x and y coordinates for the hexagon centers. We offset every other column by half the vertical spacing. This offset creates the staggered layout of the grid.

4. Hexagon Construction

We calculate the six vertices for each center using polar coordinates. We connect the vertices to create the hexagon shape.

5. Grid Assembly and Identification

We save all hexagons in a GeoDataFrame. Each hexagon has the same coordinate reference system as the input data. We give each hexagon a unique ID (cell ID).

This hexagonal grid covers the study area completely. Figure 4 is an example of these flat-topped hexagons. It provides the spatial structure for building the Chronnet network.

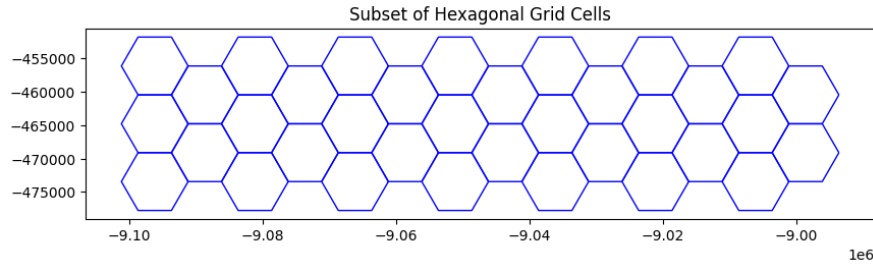


Figure 4: Flat-topped hexagons

3.3.2 Time Aggregation

To deal with the sparsity problem in Chronnet graphs, and based on the fact that satellites pass over Colombia twice a day and that fire detections are concentrated in a few hours as shown in Section 3.2, we aggregated the fire detection data into 12-hour intervals. This means that all observation times (*acq_time*) are grouped into two half-day periods: one from 00:00 to 11:59 and the other from 12:00 to 23:59. This grouping not only matches the satellite’s observation frequency but also helps maintain the correct temporal sequence of fire detections.

From the perspective of Chronnet construction, this grouping strategy is important. Chronnet builds links between fire events in temporal order to form a spatio-temporal network. Without this temporal aggregation, the fire events in Colombia would be scattered across many narrow time slices, resulting in a very sparse graph. In such cases, many cells would have no neighbors to connect, making it difficult to construct meaningful edges. This sparsity severely limits the ability to extract spread patterns or perform further network-based analysis. For example, when using a grid size of 5000 meters without time aggregation to construct a chronnet graph, the largest strongly connected component (SCC) contains only about 2% of all nodes. In addition, temporal aggregation helps to reduce the time bias introduced by the satellite’s observation schedule. Fires detected in the same pass often have similar timestamps because the satellite scans large areas sequentially within a short time window. However, these timestamps reflect the observation time not the actual ignition time of each fire, this could lead to temporal bias in the data. By grouping fire detection events into 12-hour window, we group nearby events together regardless of slight difference in observation time, so that the resulting graph have better approximates the real fire spread dynamics.

4 Method and Experimental Design

This section introduces the core methodology and experimental framework of the study, designed to construct a model that effectively simulates fire spread dynamics in Colombia and assesses its long-term risk. It begins by explaining the construction of the Chronnet network, an event-chronological, directed, and weighted graph. This network captures fire propagation patterns by mapping spatiotemporal fire event data onto a hexagonal grid and connecting fires that are sequential in time and adjacent in space. Building on this network, the study introduces a network-based Susceptible-Infected-Susceptible (SIS) epidemic model. This model treats fire spread as a contagion process on the

Chronnet network, where each grid cell (node) can be in either a "susceptible" or an "infected" (burning) state. By solving for the model's steady-state solution, the fire risk for each cell can be estimated. Finally, the section outlines the experimental design, including how an appropriate spatial resolution was selected through sensitivity analysis and how the resulting network's topological properties were characterized. A comprehensive evaluation framework is also defined. We first define the ground truth for the fire risk. Then we include metrics like Average Recall (AR) [23] and Normalized Discounted Cumulative Gain (nDCG) [28] to assess the model's ranking performance, while also using Jensen-Shannon Divergence (JSD) [29] as a calibration metric for distributional similarity. The model's performance is benchmarked against PageRank [13] to validate the effectiveness and potential advantages of the SIS model in assessing fire risk.

4.1 Chronnet Construction

In this section, we describe how we build Chronnet from spatial-temporal fire event data. Chronnet is an event-chronological, directed, and weighted graph. We start by introducing the notation and then explain each step of the process. These steps include data preparation and gridding, grouping events by time, linking events in space and time, and calculating edge weights.

Let

$$\mathcal{E} = \{e_i = (\ell_i, t_i) \mid i = 1, \dots, M\}$$

be our dataset of fire events. Each event e_i is recorded at a geographic location $\ell_i \in \mathbb{R}^2$ and a time $t_i \in \mathbb{R}_{>0}$. We start by assigning these events to a fixed hexagonal grid. Then we connect hexagon cells in chronological order when successive events occur close enough in space.

Formally, let

$$\mathcal{H} = \{h_1, \dots, h_N\}$$

be the set of hexagon cells, each with side length L . Each cell h_c has a unique index c and a geometry $\text{Geom}(h_c)$. We use a spatial assignment function:

$$c: \mathcal{E} \rightarrow \mathcal{H}, \quad e_i \mapsto c(e_i) \in \mathcal{H}.$$

We assign each event e_i to cell $c(e_i)$ if ℓ_i is inside $\text{Geom}(c(e_i))$.

Next, we divide time into intervals of length Δt (for example, 12 hours). For each event e_i , we compute its time bin:

$$t_i^* = \lfloor t_i / \Delta t \rfloor \Delta t.$$

Sorting events by t_i^* gives us groups:

$$\mathcal{G} = \{(h_i, t_i^*)\}.$$

We now build a directed graph $G = (V, E, W)$. Let $\{e_i\}_{i=1}^M$ be our fire-event observations, each assigned to hexagon cell $c_i = c(e_i)$ and binned into time slots

$$t_i^* = k_i \Delta t, \quad k_i \in \mathbb{Z}_{>0}.$$

For each integer k , define the sets of fire event in successive bins:

$$E_k = \{e_i \mid t_i^* = k \Delta t\}, \quad E_{k+1} = \{e_j \mid t_j^* = (k+1) \Delta t\}.$$

Let $x_{e_i}, x_{e_j} \in \mathbb{R}^2$ be the centroid coordinates of fire events e_i and e_j , respectively, $d_{\max}$ is defined as a threshold parameter that can be used to limit the maximum distance between links. To restrict connections to only geographically neighboring nodes, we set the distance threshold $d_{\max}$ to the maximum distance between two adjacent hexagonal cells, which is $2\sqrt{3} \cdot r$.

If

$$\|x_{e_i} - x_{e_j}\| \leq d_{\max},$$

and $i \in E_k, j \in E_{k+1}$, then let c_i and c_j be the grid cells containing events e_i and e_j . we then add a directed edge from cell $c(e_i)$ to $c(e_j)$, if this edge already exists, we increment its weight w_{ij} by one.

Repeating this for all k yields the full Chronnet graph. In addition, to remove noise and emphasize significant spatio-temporal patterns, we prune all edges with weight less than or equal to w_0 . This helps to retain only the strongest and most relevant connections between nodes. The algorithm 1 shows the Chronnet construction in algorithmic form.

In our implementation, the sliding window parameter h is set to 1, as our focus is on capturing consecutive fire events to better model the temporal spread of fires. The pruning threshold w_0 is set to 2 to remove weak temporal connections. The spatial distance threshold $d_{\max}$ is set to $2\sqrt{3} \cdot r$, which corresponds to the maximum possible distance between two adjacent hexagonal cells of side length r .

4.2 Network Based SIS Model for Fire Spread

This section explains how the Susceptible-Infected-Susceptible (SIS) epidemic model is adapted to a network framework to simulate fire spread dynamics. The model is applied to the constructed Chronnet graph, and the computationally efficient N-Intertwined Mean-Field Approximation (NIMFA) is used to estimate the burning probability of each node. The analysis focuses on two key aspects: deriving the steady-state solution, which represents long-term fire risk, and defining the epidemic threshold τ , a critical value based on the network's topology that determines if fire spread can be sustained.

Algorithm 1 Chronnet Construction with Temporal Aggregation

Require: Fire-events $\{e_i\}_{i=1}^M$, binned time $t_i^* = k_i \Delta t$, cell assignment $c_i = \sigma(e_i)$, centroids x_{e_i} , distance threshold $d_{\max}$, pruning threshold w_0

```

1:  $V \leftarrow \emptyset$ ,  $E \leftarrow \emptyset$ ,  $W \leftarrow \{\}$ 
2: for each integer  $k$  do
3:    $E_k \leftarrow \{e_i : t_i^* = k \Delta t\}$ 
4:    $E_{k+1} \leftarrow \{e_j : t_j^* = (k+1) \Delta t\}$ 
5:   for all  $(e_i, e_j) \in E_k \times E_{k+1}$  do
6:     if  $\|x_{e_i} - x_{e_j}\| \leq d_{\max}$  then
7:        $u \leftarrow c_{e_i}$ ,  $v \leftarrow c_{e_j}$ 
8:       if  $u \notin V$  then
9:          $V \leftarrow V \cup \{u\}$ 
10:      end if
11:      if  $v \notin V$  then
12:         $V \leftarrow V \cup \{v\}$ 
13:      end if
14:      if  $(u, v) \notin E$  then
15:         $E \leftarrow E \cup \{(u, v)\}$ ,  $W_{uv} \leftarrow 1$ 
16:      else
17:         $W_{uv} \leftarrow W_{uv} + 1$ 
18:      end if
19:    end if
20:  end for
21: end for
22: ▷ Prune weak edges
23: for all  $(u, v) \in E$  with  $W_{uv} \leq w_0$  do
24:   remove edge  $(u, v)$  from  $E$  and  $W$ 
25: end for
26: return  $G = (V, E, W)$ 

```

4.2.1 Overview of the SIS Framework

The classical SIS model is developed to capture the dynamic behavior of epidemic processes. It describes a contagion process in which each node changes between two states: susceptible (S) and infected (I). In a well-mixed population of fixed size N , the dynamics can be written as follows, as described by Paré et al. (2020) [10].

$$\dot{S}(t) = -\beta S(t) I(t) + \gamma I(t), \quad (1)$$

$$\dot{I}(t) = \beta S(t) I(t) - \gamma I(t), \quad (2)$$

where β is the infection rate (contacts per unit time leading to transmission) and γ is the recovery rate (return to susceptibility).

This model captures simple contagion dynamics but assumes every unit interacts equally with all others. However, real-world processes, such as fire spread, often happen on complex networks instead of assuming uniform mixing. In such cases, the spread dynamics depend on the specific connections among spatial units.

To account for these interactions, we adopt a network-based SIS modeling framework in which the spread of contagion is represented over a network topology. In our study, we use the Chronnet graph to represent the spatiotemporal structure of fire spread. Each node i refers to a hexagonal grid cell, and the directed edge weight w_{ij} refers to the pruned occurrence count of fire spread from cell j to cell i , derived from the Chronnet construction based on consecutive fire events.

For computational efficiency and tractability, we assume homogeneous infection and recovery rates across the entire network, denoted by β and γ respectively. Let $v_i(t) \in [0, 1]$ denote the probability that cell i is burning at time t . Here, a node is defined as burning at time slot t if there is at least one fire detection within this hexagon cell.

We estimate node infection probabilities using the mean-field approach rather than simulating the full stochastic SIS process, which improves computational efficiency. The N-Intertwined Mean-Field Approximation (NIMFA) is one of the most accurate node-level mean-field approximations for network-based SIS models. It assumes that the infection probability of a node is independent of its neighbors, as shown in Van Mieghem's work [30]. Under NIMFA, the homogeneous network-based SIS dynamics are expressed as:

$$\frac{dv_i}{dt} = (1 - v_i)\beta \sum_j w_{ij}v_j - \gamma v_i, \quad i = 1, \dots, n \quad (3)$$

where:

- $\beta > 0$ is the global infection rate, representing the average tendency of a burning cell ignites a neighboring unburned cell;
- $\gamma > 0$ is the global recovery rate, representing the tendency of a burning cell to extinguish the fire;
- $w_{i,j}$ is the weight from cell i to cell j .

In matrix form:

$$\dot{v} = \beta(I - \text{diag}(v))Wv - \gamma v, \quad v = (v_1, \dots, v_n)^\top$$

where:

- $\beta > 0$ is the global infection rate;
- $\gamma > 0$ is the global recovery rate;
- W is the adjacency matrix.

This homogeneous network-based SIS model enables us to capture the key spatiotemporal dynamics of fire spread while maintaining a balance between model complexity and computational efficiency.

4.2.2 Steady-State Solution of the SIS Model

In the SIS model, the steady-state infection probability v^* can be used as a reliable indicator for assessing fire risk. At steady state, $\frac{dV(t)}{dt} = 0$, It can be obtained from Eq 3:

$$\dot{v}_i = (1 - v_i)\beta \sum_j w_{ij}v_j - \gamma v_i = 0, \quad (4)$$

In this work, the steady-state solution of the continuous-time N-intertwined SIS model is found by solving a non-linear system of equations. This system can be written as $F(v) = 0$, where the i -th component is given by:

$$f_i(v) = \beta(1 - v_i)(Wv)_i - \gamma v_i = 0 \quad (5)$$

Here, $(Wv)_i = \sum_{j=1}^n W_{ij}v_j$, β is the infection rate, and γ is the recovery rate.

To solve this system, we use a numerical root-finding algorithm, specifically the hybr method from SciPy's `optimize.root` function, which is a variant of the Newton-Raphson method [31]. The core of Newton-like methods is an iterative

process that refines an initial guess $v^{(k)}$ to find a better approximation $v^{(k+1)}$. This update step is calculated by solving the linear system:

$$J(v^{(k)})\Delta v = -F(v^{(k)}) \quad (6)$$

where $\Delta v = v^{(k+1)} - v^{(k)}$ is the update step and $J(v^{(k)})$ is the Jacobian matrix of the function F evaluated at the current point $v^{(k)}$.

The Jacobian matrix, which contains all first-order partial derivatives of the system, is crucial for the efficiency and reliability of the solver. It provides the best linear approximation of the function at a given point, essentially creating a "map" of the function's local gradient that guides the solver to the root. By providing Jacobian matrix, it avoids the slow and computationally expensive process of numerically approximating the derivatives at every step.

The Jacobian matrix J for our system is an $n \times n$ matrix where each element J_{ik} is the partial derivative of the i -th function f_i with respect to the k -th variable v_k . We derive it by differentiating Equation 5 with respect to v_k :

$$J_{ik} = \frac{\partial f_i}{\partial v_k} = \frac{\partial}{\partial v_k} \left[\beta(1 - v_i) \left(\sum_{j=1}^n W_{ij}v_j \right) \right] - \frac{\partial}{\partial v_k} (\gamma v_i) \quad (7)$$

Applying the product rule to the first term gives:

$$\frac{\partial}{\partial v_k} \left[\beta(1 - v_i) \left(\sum_{j=1}^n W_{ij}v_j \right) \right] = \beta(1 - v_i)W_{ik} - \beta\delta_{ik} \left(\sum_{j=1}^n W_{ij}v_j \right) \quad (8)$$

where δ_{ik} is the Kronecker delta. The derivative of the second term is $-\gamma\delta_{ik}$. Combining these two parts gives the complete expression for the Jacobian element:

$$J_{ik} = \beta(1 - v_i)W_{ik} - \beta\delta_{ik}(Wv)_i - \gamma\delta_{ik} \quad (9)$$

This can be expressed more compactly in matrix form. Letting $u = Wv$, the full Jacobian matrix J is given by:

$$J = \beta \cdot (\text{diag}(1 - v) \cdot W - \text{diag}(u)) - \gamma \cdot I_n \quad (10)$$

where $\text{diag}(x)$ denotes a diagonal matrix with the elements of vector x on its diagonal, and I_n is the $n \times n$ identity matrix. This is the precise formula implemented in our code to ensure an efficient and robust computation of the SIS model's steady-state solution.

The following algorithm 2 shows the procedure for computing the steady-state infection rate:

Algorithm 2 Steady-State Solution of Continuous-Time N-Intertwined SIS Model

```

1: function COMPUTE SIS STEADY STATE( $W, \beta, \gamma$ )
2:    $n \leftarrow$  number of rows of  $W$ 
3:    $p^{(0)} \leftarrow (1, \dots, 1)^\top$  ▷ initial guess
4:   function  $F(p)$ 
5:      $u \leftarrow W p$ 
6:     return  $\beta \cdot ((1 - p) \odot u) - \gamma p$ 
7:   end function
8:   function  $J(p)$  ▷  $J$  is the Jacobian of  $F$ 
9:      $u \leftarrow W p$ 
10:    return  $\beta (\text{diag}(1 - p) W - \text{diag}(u)) - \gamma I_n$ 
11:  end function
12:   $v \leftarrow \text{RootSolve}(F, J, p^{(0)}, \text{method} = \text{hybr}, \text{tol} = 10^{-8}, \text{maxfev} =$ 
    1000)
13:  if not Converged( $v$ ) then
14:     $v \leftarrow (0, \dots, 0)^\top$ 
15:  end if
16:  return  $v$ 
17: end function

```

4.2.3 Fire Spread Threshold

In the stability analysis of the classic SIS model under homogeneous mixing, the **basic reproduction number** R_0 , defined as

$$R_0 = \frac{\beta}{\gamma},$$

captures the potential severity of the viral spread. The corresponding epidemic threshold is $R_0 = 1$, meaning:

- If $R_0 > 1$, the infection can spread and lead to an epidemic outbreak.
- If $R_0 \leq 1$, the infection dies out and the system converges to a disease-free state.

In contrast, in the N-intertwined mean-field approximation (NIMFA) of the network-based SIS model, the epidemic threshold depends on the underlying topology. It is given by [30]

$$\tau_c = \frac{1}{\lambda_{\max}(W)},$$

where $\lambda_{\max}(W)$ is the largest eigenvalue (i.e., spectral radius) of the adjacency matrix W of the network.

In this context, the effective infection rate is defined as $\tau = \beta/\gamma$, where β is the global infection (fire spread) rate and γ is the global recovery (extinguishing) rate, then the condition for epidemic persistence becomes:

- If $\tau > \tau_c$, or equivalently $\frac{\beta}{\gamma} > \frac{1}{\lambda_{\max}(W)}$, the infection can persist in the network and may reach a non-zero steady state.
- If $\tau \leq \tau_c$, or equivalently $\frac{\beta}{\gamma} \leq \frac{1}{\lambda_{\max}(W)}$, the infection dies out and the system tends to a disease-free equilibrium.

4.3 Experiment Design

This section introduces the experimental procedure, starting with the setup and characterization of the network, then then moving on to the implementation, calibration, and evaluation of the SIS model.

4.3.1 Experimental Setup and Network Characterization

This part outlines the preparatory stages of our experiment, from selecting the appropriate spatial resolution to characterizing the resulting network structures.

First, to understand how spatial resolution impacts the structure and spread dynamics of the chronnet graph, we performed a grid resolution sensitivity analysis. Specifically, we test a wide range of hexagonal cell sizes: 1000, 2000, 2375, 2750, 3000, 4000, 5000, 6000, and 7000 meters. For each resolution, we built a Chronnet and computed two metrics: (1) the percentage of nodes contained in the largest strongly connected component (SCC), and (2) the percentage of network-driven fire events occurring within that SCC. The first metric reflects the structural connectivity of the network at a given resolution, while the second captures the strength of potential spread dynamics inferred from event sequences. These metrics help identify grid sizes that balance network connectivity and meaningful fire propagation patterns. Here, we compute metrics focusing on SCC because it show good convergency quality for the following SIS modeling and will be studied independently. The results of this sensitivity analysis are discussed in Section 5.1.1. After the grid resolution sensitivity analysis, we select the 2000m, 3000m, and 5000m as representative cases for all subsequent modeling, as they correspond to high, medium, and low spatial resolution scenarios.

For these three selected grid size, we then performed a detailed characterization of the Chronnet’s structural properties, with a focus on the two largest SCCs. Specifically, we compute a set of structural metrics for each SCC:

- The number of nodes and edges
- The percentage of nodes and edges relative to the full graph
- The average shortest path length (L)
- The clustering coefficient (C)
- The small-worldness index S

The small-worldness index is designed to assess the extent to which each SCC exhibits small-world properties:

$$S = \frac{C/C_{\text{rand}}}{L/L_{\text{rand}}},$$

where C and L are the clustering coefficient and average shortest path length of the observed graph, and C_{rand} and L_{rand} are the corresponding values computed from a random graph with the same number of nodes and edges. A network is typically considered small-world if $S > 1$. Since NetworkX, a Python package for network analysis, does not provide a built-in small-worldness function, we implemented this computation by generating 10 random directed graphs (with fixed edge count) using the `nx.gnm_random_graph` function and computing its average clustering and path length.

In addition, we explore the relationship between basic node metrics and the ground truth burning probability v_{net} . This analysis aims to identify whether structural positions in the network can serve as naive indicators of fire vulnerability, and to what extent they correlate with actual fire occurrence patterns. These node metrics include the following:

- **Degree (k_i):** The degree of a node i is the number of direct connections (edges) it has to other nodes.

$$k_i = \sum_j A_{ij}$$

where A is the adjacency matrix of the network. A higher degree may suggest greater potential for fire spread due to more neighboring nodes.

- **Strength (s_i):** In a weighted network, the strength of a node is the sum of the weights of its connected edges.

$$s_i = \sum_j w_{ij}$$

where w_{ij} is the weight of the edge between nodes i and j . Strength captures both the number and intensity of connections.

- **PageRank (PR_i):** PageRank measures the importance of a node based on the idea that connections from high-importance nodes contribute more. It is computed iteratively as:

$$PR_i = \alpha \sum_{j \in \mathcal{N}_i} \frac{PR_j}{k_j^{\text{out}}} + (1 - \alpha) \frac{1}{N}$$

where $\mathcal{N}_i$ is the set of in-neighbors of node i , k_j^{out} is the out-degree of node j , N is the total number of nodes, and α is the damping factor (commonly 0.85).

- **Eigenvector Centrality (EC_i):** This metric assigns relative scores to nodes based on the concept that connections to high-scoring nodes contribute more to the score of the node.

$$EC_i = \frac{1}{\lambda} \sum_j A_{ij} EC_j$$

where λ is a constant (the largest eigenvalue of the adjacency matrix A). Nodes connected to influential nodes receive higher scores.

- **Betweenness Centrality (BC_i):** Betweenness centrality measures how often a node lies on the shortest paths between other pairs of nodes.

$$BC_i = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from node s to node t , and $\sigma_{st}(i)$ is the number of those paths that pass through i . High betweenness may indicate a bottleneck node in fire propagation.

- **Closeness Centrality (CC_i):** Closeness centrality of a node i is the reciprocal of the average shortest-path distance to all other nodes that are reachable from i . This metric reflects how efficiently a node can access other

parts of the network. A higher value indicates that a node is, on average, closer to all other nodes, allowing it to spread influence (or fire) more rapidly.

$$CC_i = \frac{n_i - 1}{\sum_{j \in R_i, j \neq i} d(i, j)}$$

where R_i is the set of nodes reachable from node i , and $n_i = |R_i|$ is the number of reachable nodes. If a node is isolated ($n_i = 1$), its closeness centrality is defined as 0. The term $d(i, j)$ represents the shortest-path distance from i to j .

- **Clustering Coefficient (C_i):** This metric quantifies how close the neighbors of a node are to being a complete clique (i.e., fully connected among themselves):

$$C_i = \frac{2T_i}{k_i(k_i - 1)}$$

where T_i is the number of triangles (fully connected triplets) involving node i , and k_i is the degree of node i . A high clustering coefficient may imply more localized spread potential.

4.3.2 Modeling and Evaluation Framework

This part introduces our framework for modeling fire risk using the SIS model and for evaluating its performance.

A core component of our evaluation is a well-defined **ground truth**. Based on the observation that network-driven fires show distinct spatio-temporal patterns compared to spontaneous fires (see 5.1.2), and given that Chronnet inherently captures only network-driven fire propagation, we label fire events as either spontaneous or network-driven (self-loops are considered as network-driven). Then we define the overall burning probability as a hybrid target:

$$v_{\text{hybrid}} = v_{\text{sf}} + (1 - v_{\text{sf}})v_{\text{net}},$$

where

- v_{hybrid} is the overall burning probability of each cell,
- v_{sf} is the spontaneous fire probability, and
- v_{net} is the conditional probability of a network-driven fire, given no spontaneous fire.

Here, a cell is considered to be burning if at least one fire event occurs within it. Specifically, cell i is labeled as having a spontaneous fire at time t if none of its neighbors had a fire event at time $t - 12\text{h}$. Similarly, a network-driven fire at cell i and time t is defined as a fire event where at least one neighboring cell was burning at $t - 12\text{h}$. This condition aligns with the criteria for link creation in the Chronnet construction method.

We estimate the probabilities of these events from historical fire records. Let $N_{burn,i}$ be the total number of burning occurrences in cell i , $N_{sf,i}$ the number of spontaneous burning events, and $N_{obs,i}$ the total number of observations (i.e., time bins). The overall burning probability ($v_{hybrid,i}$) and the spontaneous fire probability ($v_{sf,i}$) are:

$$v_{hybrid,i} = \frac{N_{burn,i}}{N_{obs,i}}$$

$$v_{sf,i} = \frac{N_{sf,i}}{N_{obs,i}}$$

Since the SIS model is designed to simulate the spread dynamics within the network, its estimated steady-state infection probability should correspond to the empirical probability of network-driven fires. We define this as our ground truth, v_{net} , and derive it as follows:

$$v_{net,i} = \frac{N_{burn,i} - N_{sf,i}}{N_{obs,i} - N_{sf,i}}$$

This value, $v_{net,i}$, represents the conditional probability of a fire occurring in cell i due to network spread, given that it was not a spontaneous ignition. It will serve as the target for evaluating our model's performance. This definition includes both fire persistence (self-loops) and inter-cell spread, since initial data exploration in Section 5.1.3 showed that fire persistence with the same cell is an important part of fire activity. To check that this choice does not bias the evaluation, we also conduct a validation analysis in Section 5.1.3 to confirm that the cell risk rankings are highly correlated with or without self-loops. This supports using the broader definition of v_{net} as a stable ground truth for the main evaluation.

To measure the model's performance, we use three **evaluation metrics**, each capturing a different aspect of performance:

- **Average Recall (AR):** This metric evaluates the **ranking quality** of the model's estimation. It measures how well the estimated ranking of burning probability aligns with the ground truth ranking. It is defined as:

$$AR = \frac{1}{N} \sum_{k=1}^N \frac{|Top - k_{pred} \cap Top - k_{true}|}{k}$$

where $Top - k_{pred}$ and $Top - k_{true}$ are the sets of top-k ranked cells according to the estimated scores and the true v_{net} values, respectively. A higher AR indicates better overall ranking agreement.

- **Jensen-Shannon Divergence (JSD):** This metric serves as a **calibration metric**. It measures the similarity between the distribution of estimated steady-state probabilities and the distribution of the ground-truth v_{net} values. Both distributions are estimated using histograms with $B = 100$ bins over the overlapping range of positive values. Before computing the JSD, we scale the values to the $[0,1]$ interval using min-max normalization. This normalization is crucial because the two distributions operate on different scales. The empirical ground truth probabilities are very low (in the $[0, 0.1]$ range), while the SIS model's estimation can span to a much wider range that can close to 1.0, particularly when using a high effective infection rate τ . The experiments with high τ are necessary because our objective is to optimize the model's ranking performance, rather than achieving precise probabilistic estimation. However, a good ranking metric alone doesn't guarantee that the assigned scores carry real-world meaning. For example, when the τ becomes high, the SIS model tends to saturate: most nodes end up with similar value near 1, making the estimation nearly uniform. This setting can still achieve high Recall (AR) scores, but the output loses its ability to differentiate between high and low risk areas. Given this setting, it is important to ensure that the estimated probability distribution still preserves meaningful relative differences and is not saturated. So we use JSD as a calibration metrics to make that the distribution of estimated probabilities should resemble the observed distribution of network-driven burning probability.
- **Normalized Discounted Cumulative Gain (nDCG):** To more evaluate the model's ability to prioritize the most critical high-risk cells, we use nDCG. Unlike AR, nDCG places more weight on the correct ranking of top-tier items, which is crucial for practical applications like resource allocation in fire prevention. The nDCG at rank k is calculated as:

$$nDCG@k = \frac{DCG@k}{IDCG@k} \quad \text{where} \quad DCG@k = \sum_{i=1}^k \frac{rel_i}{\log_2(i+1)}$$

Here, rel_i is the relevance score of the item at position i (in our case, the v_{net} value), and IDCG is the DCG score of a perfect ranking. We will compare the nDCG@k scores of the SIS model and PageRank for different values of k (e.g., 100, 300, 500, 1000).

Regarding the **SIS implementation part**, our goal is to compute the steady-state infection probability, which we use as a proxy for the estimated burning probability of each node. For each grid size, we apply the SIS model to the two largest strongly connected components (SCCs) respectively obtained from the Chronnet graph. In our experiments, we first normalize the adjacency matrix W by dividing it by its largest eigenvalue $\lambda_{\max}(W)$. This transforms the epidemic threshold for each SCC from $1/\lambda_{\max}(W)$ to 1, allowing us to test various values of $\tau = \beta/\gamma$ across different graphs under the same threshold scale. We fix the recovery rate $\gamma = 1$, without losing generality, the infection rate β is inferred as $\beta = \tau \cdot \gamma$. To compute the steady-state infection probabilities, we solve the equations using the `scipy.optimize.root`.

The performance of the SIS model is highly dependent on the effective infection rate, $\tau = \beta/\gamma$. To **select the optimal** τ for each SCC obtained from chronnet graph, we perform a parameter sweep and evaluate the model’s output against the ground truth v_{net} using two distinct metrics: Average Recall (AR) and Jensen-Shannon Divergence (JSD). In our experiment, we conducting a parameter sweep over 100 values ranging from 1.1 to 100. The steady-state infection probabilities are computed for each τ using `scipy.optimize.root`. The optimal τ is then selected using a systematic procedure that aims to find a trade-off point that minimizes JSD while maximizing a high and stable AR, so that the optimal τ ensure both strong ranking performance and interpretable risk scores. In practice, for each (grid size, SCC rank) combination, we selected the τ value at the first turning point where JSD changes from decreasing to increasing, and AR has already stabilized, which we define as the first point where AR changes by less than 0.002 across two successive τ values. This turning point captures the trade-off between ranking stability and distribution fidelity. In cases where no such turning point was observed, we selected the τ with the lowest JSD among those on the $\geq 95\%$ AR plateau, ensuring both high ranking performance and low divergence. Finally, we applied a post-selection check: if any other τ yielded both a lower JSD and a higher AR than the initially selected one, we replaced it with the τ that achieved the highest AR among these strictly better alternatives. This final step ensures that no dominated solution is retained as optimal.

Once the optimal τ is determined, we conduct the final performance evaluation. The calibrated SIS model’s estimation are compared against a PageRank baseline. This comparison helps determine if the explicit modeling of epidemic dynamics provides additional explanatory power. The final comparison uses all three metrics (AR, JSD, and nDCG) to provide a comprehensive assessment of whether the dynamic, non-linear SIS model offers an advantage over a static, centrality-based approach in identifying fire risk. By using these three metrics, we evaluate not only its overall ranking performance but also its specific

effectiveness in identifying top-priority areas and its ability to reproduce the observed spatial risk distribution.

Here, we chose PageRank as the baseline because it offers a good balance between theoretical comparability and practical interpretability. Both PageRank and the SIS model describe a kind of steady-state behavior. PageRank reflects the long-term visit probability in a damped random walk, while the SIS model captures the long-term infection probability under a mean-field approximation. This makes them conceptually comparable. Besides, PageRank accounts for the influence of high-risk neighbors, which better aligns with how fire spreads through vulnerable areas. It's also computationally efficient and easy to reproduce, which is useful for large-scale evaluations. More importantly, if the SIS model consistently outperforms PageRank, it suggests incorporating actual epidemic dynamics provides additional explanatory power beyond what static graph structure can offer. Other baselines like degree and strength were excluded. We did not select degree as a baseline because it lacks sufficient discriminative ability in our network. As shown in Figure 7, the degree distributions are highly left-skewed, with over 80% of nodes having a degree less than 10. This leads to frequent overlap among node degrees, making it difficult to distinguish node importance effectively. Strength was also excluded because it directly reflects the frequency of network-driven fire events, which are already used in the construction of our ground truth, so using strength would lead to a circular evaluation and cannot provide a fair comparison for model performance.

Finally, to separate the effects of inter-cell spread from intra-cell persistence, we conduct an ablation study to assess the impact of self-loops in the SIS model. In this study, the SIS model was established on a modified Chronnet graph where all self-loop edges were removed. By comparing the performance of full network against the no-self-loop network, we can determine whether the model's explanatory power stems from capturing the spread between cells or if it relies heavily on the signal from fire persistence. This analysis provides deeper insight into the robustness of the modeling framework and the nature of the dynamics it captures.

5 Results and Analysis

This section presents a comprehensive analysis of the experimental results, beginning with a validation of the modeling approach. We first investigate how spatial resolution influences the inferred fire spread patterns in the Chronnet, a step that justifies the selection of 2000m, 3000m, and 5000m as representative grid sizes for subsequent analysis. Next, we validate the ground truth defi-

dition, showing that it is reasonable to distinguish between spontaneous and network-driven fires, and that it is robust to including self-loops. Then the section characterizes the topological properties of these networks, examining their degree and strength distributions, small-world characteristics, and the relationship between various node metrics and observed fire behavior. The core of the section evaluates the performance of the network-based SIS model. We systematically compare its ability of ranking high-risk cells against a PageRank baseline using three metrics: Average Recall (AR), Jensen-Shannon Divergence (JSD), and nDCG. In addition, a qualitative assessment is provided by comparing the spatial distribution of the model’s estimated long-term fire risk with the ground truth patterns, confirming the model’s utility in identifying high-priority regions. Finally, we conduct an ablation study to assess the impact of self-loops on performance.

5.1 Method Validation

5.1.1 Chronnet Resolution Effects on fire Spread Patterns

In order to understand how spatial resolution influences the inferred spread of fire in Chronnet, and to justify the selection of representative grid sizes for subsequent analysis, we performed a sensitivity analysis by varying the grid cell size between 1000 m and 7000 m. For each grid size, we built a Chronnet graph. Each node represents a grid cell. Directed edges link consecutive events in adjacent cells within a fixed time window of 12 hours. We then computed two metrics: (1) the percentage of nodes contained in the largest strongly connected component (SCC), and (2) the percentage of “network-driven” fire events in that SCC. The percentage of nodes in the largest SCC reflects the structural connectivity of the spatial network at a given grid size, and the percentage of network-driven fires in that SCC indicates the strength of spread dynamics at that resolution.

Here, “Network-driven” fire events for node i are defined as events occurring at time t that had at least one neighboring node burning at time $t-12h$. This enable us to distinguish between spontaneous and network-driven fires. Since Chronnet is built from event sequences, it can only capture network-driven spread.

Fig 5 show the result. The figure shows a clear U-shaped, resolution-dependent pattern. At high resolution (1,000–2,375 m), the largest SCC covers less than 25% of all active cells. As grid size increases, the largest-SCC node percentage rises, but the network-driven fire events percentage falls, reaching a minimum at around 3,000 m. This suggests that, as grid size increases, the largest SCC begins to absorb many sparsely active nodes with few network-driven fire

events, reducing the density of network-driven fire events. However, when grid size exceeds 3,000 m, the largest-SCC percentage increase more slowly, indicating that further increases in grid size gives smaller connectivity gains; In particular, at low resolution (grid size $> 5,000$ m), the largest-SCC percentage remains almost constant, but the network-driven event percentage increase sharply. This means more and more fire events have at least one “neighbor” burning within the previous 12 hours. Because the grid is so large, cells that are physically too far apart are treated as adjacent. As a result, many truly independent fires mistakenly are counted as network-driven, leading to a higher density of network-driven fires than in the ground truth.

In addition, even at the largest grid size, fewer than one-third of the events are actually connected through fire spread, indicating that the raw dataset are dominated by spontaneous fire. Since spontaneous fires cannot be learned by Chronnet or the SIS model, we should distinguish between network-driven and spontaneous fire events when evaluating model performance.

Based on this sensitivity analysis, we selected three representative grid sizes for our main experiments: 2000 m, 3000 m, and 5000 m. This selection was not arbitrary but was made to capture the distinct network topologies observed across the resolution spectrum. The 2000 m grid represents a high-resolution, fragmented network ideal for studying local dynamics. The 3000 m grid offers a balanced middle ground between the high and low resolutions. Finally, the 5000 m grid is a low-resolution, highly aggregated network. It allows us to evaluate the model’s performance on large-scale patterns, even with some potential noise. Analyzing these three distinct cases ensures a comprehensive evaluation of how network structure, as influenced by spatial scale, affects the SIS model’s performance and outcomes

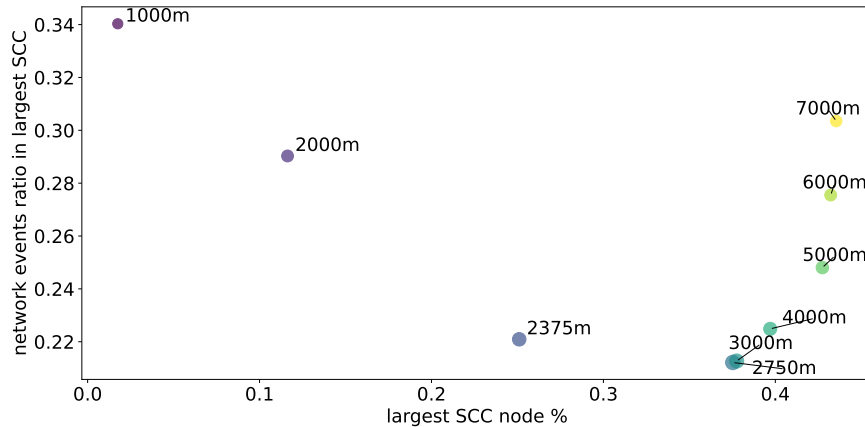


Figure 5: Largest SCC Ratio vs. Ratio of Network Events in Largest SCC

5.1.2 Distinguishing Spontaneous and Network-Driven Fire Patterns

Since previous section showed that many fires in Colombia are isolated events, not part of a clear spreading chain, it is essential to distinguish spontaneous and network-Driven fire types. The distinction is important because chronnet and SIS model is designed to simulate spread dynamics within the network, not the spontaneous ignitions. To investigate the relationship between these two fire types, we calculated the Spearman correlation between the historical occurrence counts of spontaneous and network-driven fires for each cell.

The results, presented in the table 2, show a weak to moderate correlation. This correlation becomes weaker as the grid size gets smaller. At a small grid size like 2000m, the high precision shows a very weak correlation ($\rho=0.188$), indicating the locations of spontaneous fire are largely different from the areas where fires persistently spread through the network. As the grid size increases to 5000m, an aggregation effect slightly increases this correlation to a moderate level ($\rho=0.526$).

To visually complement this finding, we also plot the geographic distributions of the actual risk for both fire types, as shown in Figure 6. For this visualization, we only use the 5000m grid size because it offers a clear balance between spatial detail and readability.

In summary, because the correlation is never strong, the analysis confirms that spontaneous and network-driven fires tend to occur in different areas, and this difference is most obvious at smaller grid sizes. This finding validates our methodological choice to evaluate the SIS model’s performance specifically against the probability of network-driven fires v_{net} .

Grid Size	Spearman’s ρ
2000	0.188
3000	0.318
5000	0.526

Table 2: Spearman correlation coefficients between Spontaneous and Network-Driven Fire Occurrence for different grid sizes (all p values < 0.001).

5.1.3 Robustness of the Ground Truth Definition

Our choice to use a broader definition of v_{net} that includes self-loops is driven by the high frequency of fire persistence observed in the data. As shown in Table 3, the proportion of fires persisting within the same cell (self-loop fires)

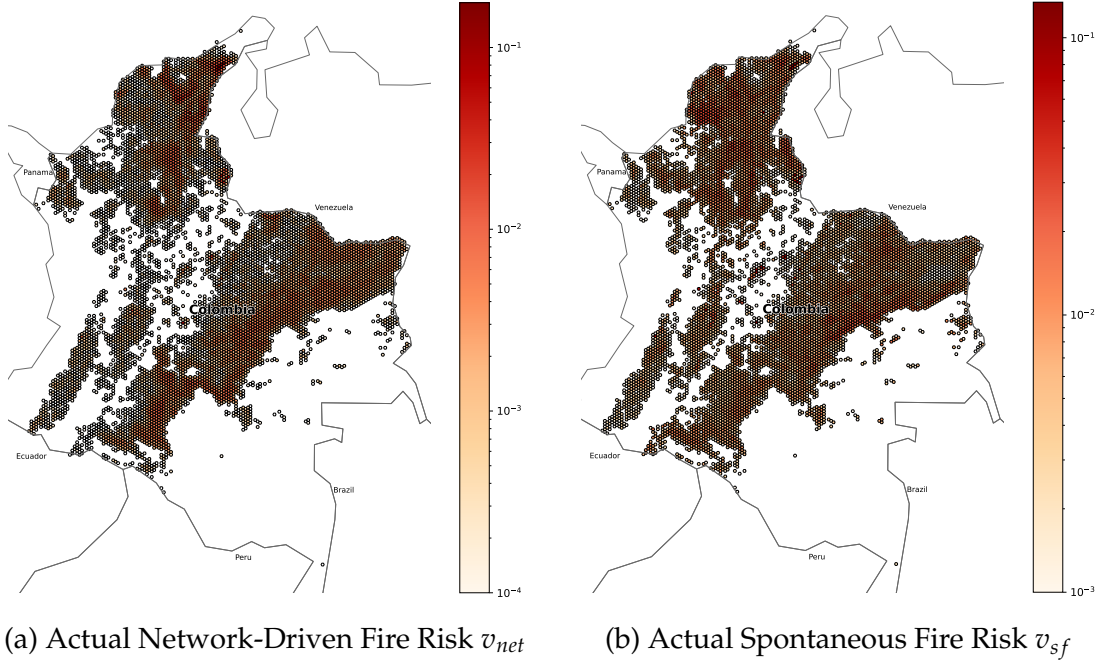


Figure 6: Spatial Distribution Comparison between Spontaneous and Network-Driven Fire Risk at 5000m Grid Size

is often comparable to that of fires spreading between adjacent cells (inter-cell fires).

One might question the inclusion of self loops (representing fire persistence) in the definition of ‘network-driven’ fire, as this term often implies spread to neighbor cells, and this inclusion could distort the risk ranking obtained purely from inter-cell spread. To address this, we computed the network driven burning probability twice: once including self loops ($v_{netw, self_loops}$) and once excluding them ($v_{netwo, self_loops}$), and then computed the spearman correlation between them.

Table 3 showed that there is very high spearman’s ρ between these two metrics. This high correlation shows that the ranking of locations by their network-driven risk remains highly stable, regardless of whether fire persistence (self-loops) is included in the calculation. This suggest that cells with high susceptibility to inter-cell spread also tend to show high persistence (self-loops). This finding validates our choice to use the broader definition of v_{net} for evaluation, as it represents a measure of a cell’s overall non-spontaneous fire activity.

Grid Size	Spearman's ρ (with vs. without self-loops)	Spontaneous fire %	Self-loop fire %	Network (inter-cell fire) %
2000	0.83	84.9%	8.3%	6.7%
3000	0.89	84.2%	7.7%	8.1%
5000	0.96	79.4%	7.8%	12.8%

Table 3: Fire event composition and spearman's ρ between v_{net} computed with and without self-loops

5.2 Chronnet Network Characterization

This section introduces structural properties of the Chronnet graphs generated at the three selected resolutions.

5.2.1 Degree and Strength Distributions of Chronnet

Figure 7 shows the degree distributions for the entire chronnet network and its two largest (SCC 1 and SCC 2) at three grid resolutions (hexagon side length $r = 2000, 3000, 5000$). Each subplot shows the percentage of nodes at a given degree k . The entire chronnet network is shown on the left, SCC 1 in the center, and SCC 2 on the right. Figure 8 shows the same distributions on log-log scatter plots. These plots helps to understand the distribution of the fire events and how the tail of the distribution behaves.

At all resolution, the degree distribution show heavy tail pattern in the entire chronnet graph. This means that that most nodes have less degree and only a few nodes have high degree. These high-degree hubs is highly responsible for spreading the fire across distant areas. In SCC 1 and SCC 2, the degree values are more balanced, this is expected because SCC are more connected than the full network. When the grid size increases from 2000 to 5000, the tail in all plots become heavier, this suggest that more nodes are linked and the network tend to become more dense.

When comparing SCC 1 with SCC 2, Figure 7 shows that SCC 1 has a broader peak, particularly at larger grid size. In the 3000 and 5000 cases, SCC 1 show a shoulder or secondary peak, while SCC 2 is mostly unimodal. This suggests that SCC 1 has a more dispersed degree distribution and greater structural heterogeneity.

Figure 9 shows the log-log strength distribution for the entire chronnet network and its two largest (SCC 1 and SCC 2) at three grid resolutions (hexagon side length $r = 2000, 3000, 5000$). In the entire chronnet network, the linear trend suggests an underlying power-law pattern.

5 RESULTS AND ANALYSIS

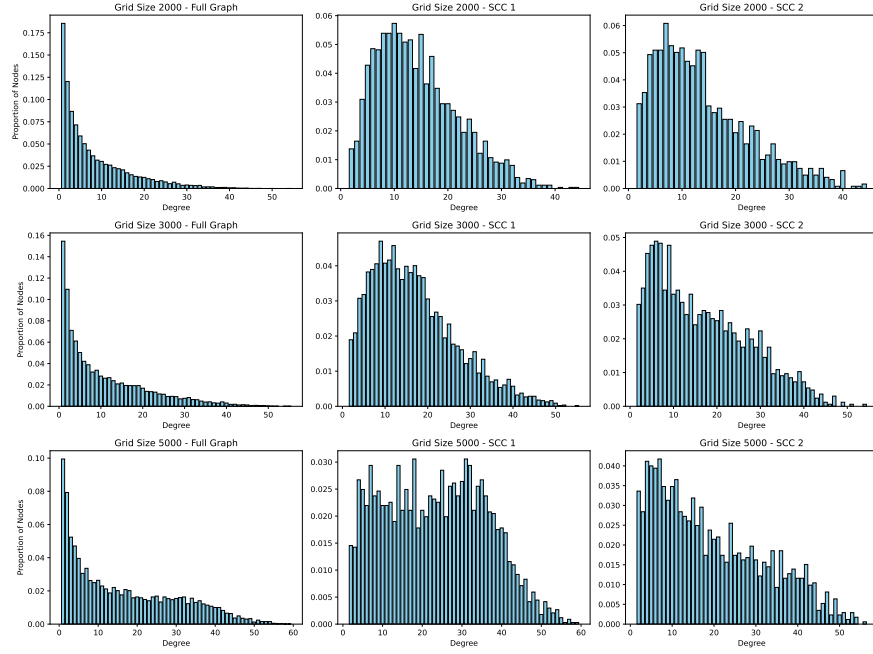


Figure 7: Degree Distribution (Bar Plot)

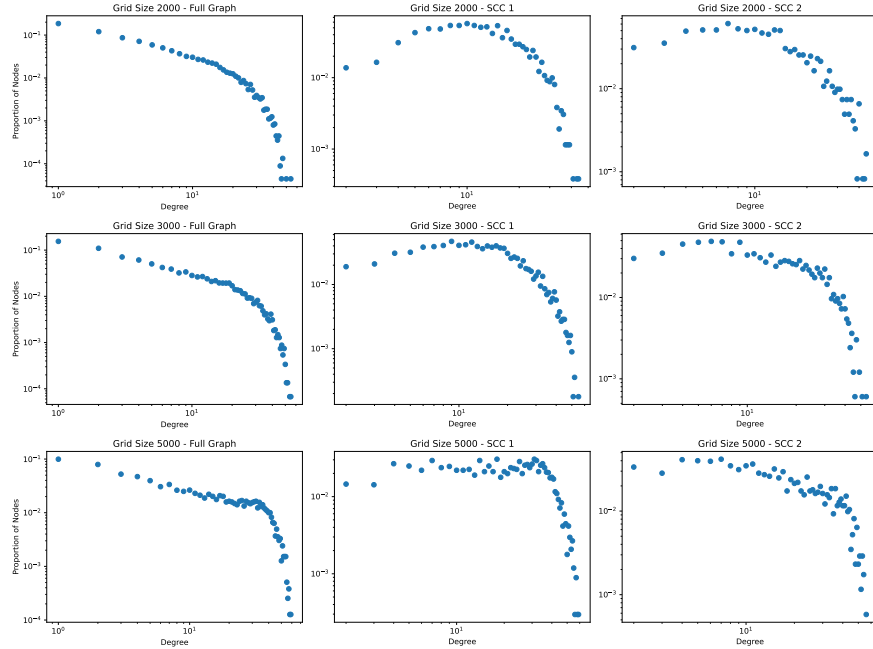


Figure 8: Degree Distribution (Log-Log Plot)

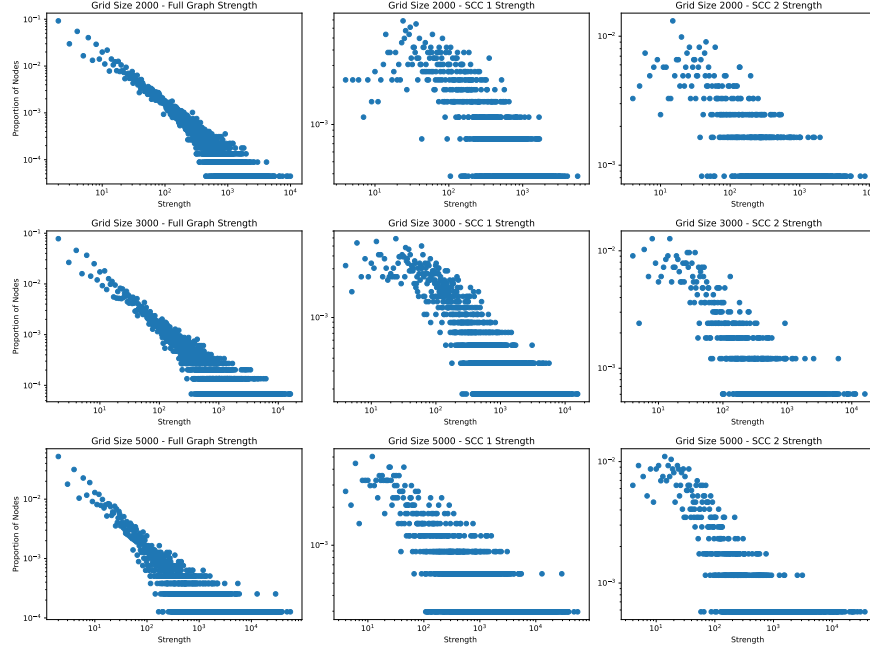


Figure 9: Strength Distribution (Log-Log Plot)

5.2.2 Topological Metrics of Top 2 SCCs

Table 4 shows the topological metrics for the two largest SCCs in Chronnet at different grid sizes. As the hexagon side length r increases from 2,000 to 5,000 meters, both of the two largest strongly connected components (SCCs) in Chronnet grow in terms of size and density. For SCC 1, the proportion of nodes increases from 11.6% to 42.7%, and edges from 20.6% to 62.2%. SCC 2 also grows from 5.4% to 21.9% of nodes and from 9.2% to 26.3% of edges. The average degree rises steadily as well (SCC 1: $7.16 \rightarrow 8.43 \rightarrow 11.75$; SCC 2: $6.92 \rightarrow 8.38 \rightarrow 9.68$), which shows that the networks become denser with larger grid sizes. This might be because more fire events fall into each cell, creating more links between them.

All SCCs have high clustering coefficients C (around 0.43–0.46), which means that the nodes tend to form tight local clusters, this is expected since nearby regions typically share similar climate condition and land use. At the same time, the networks also have small average path lengths (L from about 14.9 to 43.2), the presence of hubs play a significant role in reducing the average path length. As a result, the small-worldness values (between 12 and 29) indicate that these SCCs are small-world networks [32] [33].

Compared with SCC 2, SCC 1 contains more nodes, a denser edge structure and consistently higher small-worldness at all resolution.

Grid Size	SCC Rank	nodes	edges	nodes (%)	edges (%)	avg-deg	L	C	small_worldness
2,000	1	2,617	18,725	11.6	20.6	7.16	28.29	0.44	23.68
	2	1,217	8,422	5.4	9.2	6.92	21.69	0.43	13.26
3,000	1	5,595	47,155	37.8	60.9	8.43	43.16	0.43	28.66
	2	1,656	13,869	11.2	17.9	8.38	26.99	0.44	12.18
5,000	1	3,369	39,570	42.7	62.2	11.75	23.09	0.46	20.23
	2	1,725	16,709	21.9	26.3	9.68	14.90	0.44	18.43

Table 4: Network metrics for the two largest SCCs in Chronnet across different grid sizes

5.2.3 Relation between node metrics and fire behavior

Before we implement the SIS model to simulate fire spread dynamics, we explore how network topology relates to fire behavior, we measured the correlation between basic metrics and the ground truth burning probability. The ground truth here refers to the network-driven fire probability v_{net} described Section 4.3.2. This analysis aims to identify whether structural positions in the network can serve as naive indicators of fire vulnerability, and to what extent they correlate with actual fire occurrence patterns.

In graph theory, centrality measures are commonly used to quantify the importance or influence of nodes based on specific structural criteria. In our case, we focus on the following node-level metrics: degree, strength, PageRank, eigenvector centrality, betweenness centrality, closeness centrality and clustering coefficient. We compute the Spearman correlation between each metric and the ground truth burning probability for each SCC across three grid sizes: 2000 m, 3000 m, and 5000 m. To calculate node metrics, we used standard centrality measures. Specifically, we calculate clustering based on an undirected version of the subgraph. PageRank was calculated with a damping factor of 0.85.

Figure 14 shows the scatter plots, and the table 5 presents the corresponding Spearman correlation coefficients. Degree and strength are highly correlated with the observed fire probabilities across all grid size SCCs across all SCCs and all grid sizes, with Spearman coefficients ranging from 0.7 to 0.9. This is expected as both metrics reflect how often a node is involved in fire transitions. Since Chronnet is constructed based on sequential fire events between spatial cells, these two metrics naturally capture the intensity and frequency of network-driven fire activity.

PageRank shows a moderate to strong correlation with v_{net} , with Spearman coefficients ranging from 0.5 to 0.7. This is reasonable because PageRank

considers both the quantity and importance of incoming connections. Nodes that frequently receive fire flows from influential neighbors are assigned higher PageRank scores, which naturally align with a higher long-term likelihood of burning.

We can observe that most centrality metrics that consider the impact of far-away nodes, such as eigenvector, closeness, and betweenness, are not effective for estimating the network-driven fire probability, this is shown by their low correlation with the ground truth values. It suggests that fire dynamics in Chronnet are mainly governed by local, short-range interactions rather than by global network position. In other words, a cell’s risk depends more on the frequency of direct inflows from its neighbors, rather than its accessibility from distant parts of the graph.

Clustering coefficient also shows weak correlation with network driven burning probability. This is expected because clustering measures how well a node’s neighbors are connected to each other, while fire risk in Chronnet is driven mainly by the incoming weighted flow to the node and by the direction and temporal order of spread. In our computation, the clustering coefficient is based on an undirected version of the graph, so it ignores both edge direction and weights. Many high-risk cells receive strong incoming influence from multiple sources that are not connected to each other, which yields low clustering but high risk.

Grid Size	SCC Rank	degree	strength	pagerank	eigenvector	betweenness	closeness	clustering
2,000	1	0.74	0.69	0.54	0.25	0.43	0.19	-0.12
	2	0.86	0.83	0.62	0.43	0.43	0.43	-0.02
3,000	1	0.88	0.80	0.63	0.34	0.49	0.02	0.02
	2	0.89	0.89	0.70	0.36	0.52	0.16	0.02
5,000	1	0.92	0.85	0.62	0.41	0.52	0.28	0.09
	2	0.90	0.91	0.66	0.49	0.46	0.32	0.09

Table 5: Spearman correlation (r) between Node Centrality and v_{net} for different grid sizes SCC

5.3 SIS Model Performance Analysis

This section presents the core evaluation of the network-based SIS model’s ability to assess fire risk.

5.3.1 Experimental Results Across Different τ

We evaluate the performance of the SIS model compared to PageRank across different grid sizes and SCC ranks, focusing on how the choice of the effective infection rate τ influences results. The experimental results are summarized in Figure 10 and Table 6.

Figure 10 shows the experimental results across different values of τ , ranging from 1.1 to 100. Each subplot shows the performance of SIS-based node ranking under varying infection rate τ , evaluated using Average Recall (AR) and Jensen-Shannon Divergence (JSD). Points represent different τ values, colored by τ magnitude. The red star in Figure 10 marks the automatically selected optimal τ , balancing high and stable AR with low JSD. Dashed lines indicate the baseline performance using PageRank. If we aim to improve the ranking probability of the SIS model, the maximal AR achieved by the SIS model is higher than that of the PageRank baseline. This suggests that the SIS model has better ranking capability than PageRank. However, if we aim to reproduce the distribution of v_{net} , the situation is quite different. SIS achieves lower JSD than PageRank in only two cases, at grid sizes 3000 and 5000, both corresponding to the second-largest SCC. This indicates that the SIS model is not consistently good at reproducing the overall distribution. In addition, lower JSD values do not necessarily correspond to higher AR. This suggests a trade-off between accurately approximating the steady-state distribution and maximizing node ranking performance.

Across all subplots, we observe a consistent pattern: as τ increases, Average Recall (AR) rise rapidly at first and then tend to plateau beyond a certain point, while JSD improves (decreases) initially but later starts to get worse. This pattern indicates that in the moderate τ region, the steady state infection rate v_{sis} distribution becomes polarized, which improves the ranking quality captured by AR. However, when τ becomes very large, further increase brings very little change to the ranking, but the steady state infected rates of all nodes become saturated (close to 1), this mismatch with the ground truth causes JSD to deteriorate.

To select a suitable τ , we aim to find a balance point where AR is already high and stable, and JSD remains low. From Figure 10, we can see that this usually happens near the turning point where AR starts to plateau and JSD starts to rise again. These observations validate our τ -selection strategy, which identifies the trade-off point where AR stabilizes and JSD begins to deteriorate. While the actual selection is performed automatically (see Section 4.3.2), the turning points correspond well to visually observable transitions in all SCCs across grid sizes, suggesting that the selection is robust, consistent, and interpretable.

Moreover, we observe that as the grid size increases, the maximum AR

achieved by the SIS model also improves. Because larger grid cells aggregate more individual fire events, leading to a denser and more connected graph. This increased connectivity can provide more robust and continuous pathways for the SIS model to learn the spread dynamics. In addition, across all grid size, the SIS model’s best Average Recall (AR) on the second-largest SCC (SCC 2) consistently exceeds that on the largest SCC 1. This performance gap is likely stems from SCC 2 reflecting more consistent fire spread patterns, which the SIS model is better at capturing. In contrast, SCC 1, although larger, contains more varied and potentially noisy dynamics, as indicated by the degree distribution in Figure 8, which make it harder to model accurately. SCC 2 tends to represent clearer and more stable spatiotemporal pathways, which are more easily picked up by the Chronnet edges and thus offer a cleaner signal for the SIS model to learn from.

The optimal τ are showed in Table 6, it reveal that the optimal τ value is not constant but varies across different grid sizes and strongly connected components (SCCs). A notable finding for SCC 1 is the clear inverse pattern between the proportion of network-driven events and the optimal τ across grid sizes. The proportion of network-driven events (in descending order) is highest at grid size = 2000, then 5000, and lowest at 3000. In contrast, the optimal τ for SCC 1 is highest at grid size = 3000, then 5000, and lowest at 2000. This pattern suggests that τ adapts to the strength of the fire spread signals captured by Chronnet at different resolutions. When the proportion of network-driven events is low (e.g., grid size = 3000), implying weaker or less distinct spread signals, a larger τ is needed. This higher effective infection rate pushes the SIS model to spread more aggressively, helping it overcome sparse or less coherent connectivity and better identify high-risk cells. When the proportion is high (e.g., grid size = 2000), meaning the spread dynamics are captured more efficiently, a smaller τ is enough to reach good performance without oversaturating the network. Overall, the optimal τ is not fixed but adjusts to balance propagation and differentiation depending on the network structure.

After fixing optimal τ , we compare the SIS model with PageRank across different grid sizes and SCC ranks. As shown in Table 6, the SIS model consistently achieves higher Average Recall (AR), especially in the second-largest SCCs. This suggests that the SIS model is more effective at capturing dynamically influential nodes. The superiority of the SIS also suggest incorporating actual epidemic dynamics provides additional explanatory power beyond what static graph structure can offer. However, when comparing the JSD, the SIS does not consistently outperform PageRank in approximating the distribution of fire occurrence. It achieves lower JSD than PageRank only in two cases. This supports our earlier observation: a lower JSD does not always align with higher Average Recall. It highlights a trade-off between reproducing the steady-state

5 RESULTS AND ANALYSIS

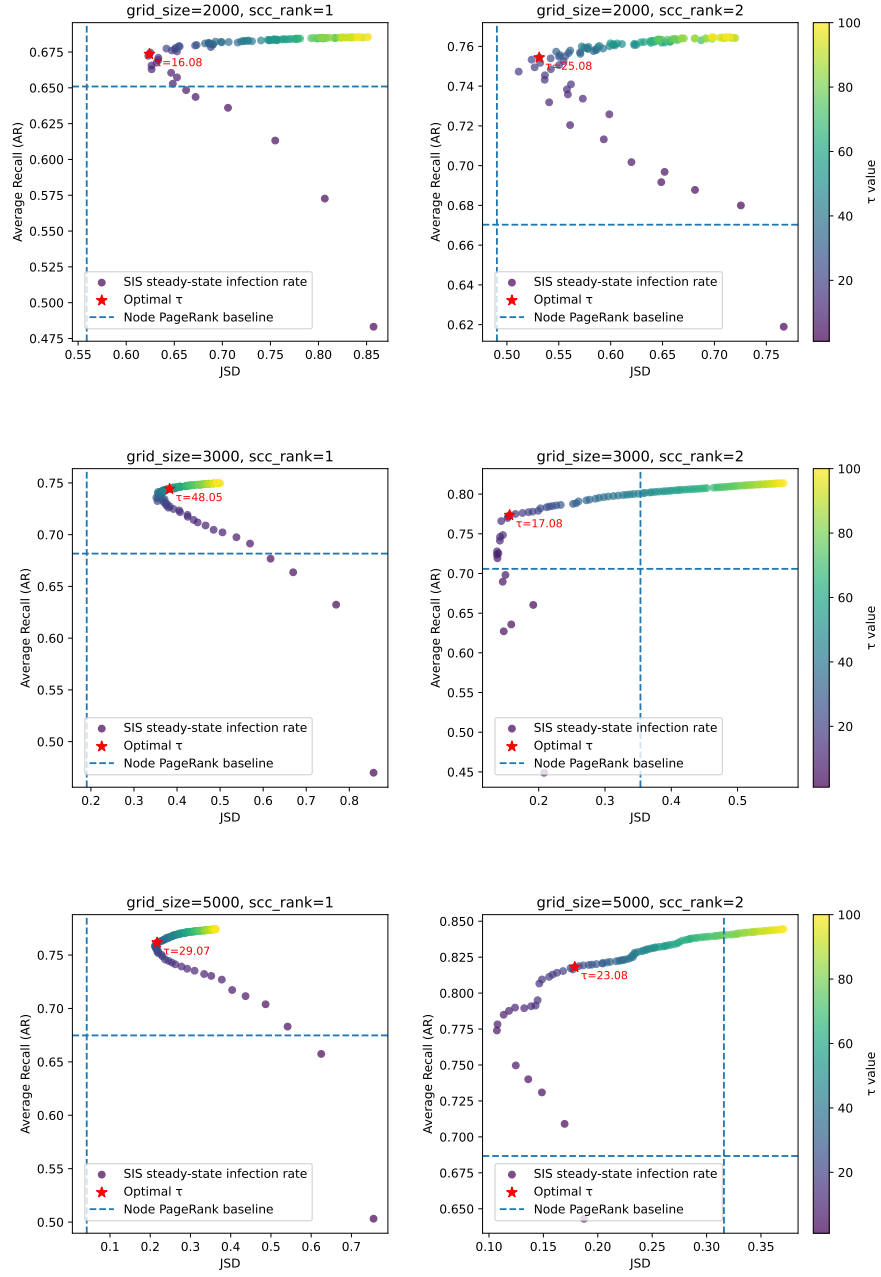


Figure 10: Trade-off between ranking accuracy and distribution similarity across τ values for different grid sizes and SCC ranks.

distribution and accurately ranking nodes by their importance. One possible reason is that we use the same β and γ for all nodes, which may limit the model’s ability to fully match the spatial heterogeneity in the empirical fire patterns.

Grid Size	SCC Rank	Optimal τ	Average Recall (AR)			JSD		
			SIS	PageRank	Δ	SIS	PageRank	Δ
2000	1	16	0.67	0.65	+0.02	0.62	0.56	+0.06
2000	2	25	0.75	0.67	+0.08	0.53	0.49	+0.04
3000	1	48	0.74	0.68	+0.06	0.38	0.19	+0.19
3000	2	17	0.77	0.71	+0.06	0.16	0.35	-0.19
5000	1	29	0.76	0.67	+0.09	0.22	0.04	+0.18
5000	2	23	0.82	0.69	+0.13	0.18	0.32	-0.14

Table 6: Comparison of Average Recall (AR) and JSD between SIS and PageRank with Optimal τ across different grid sizes and SCC ranks

5.3.2 Ranking Performance Evaluation using nDCG

To evaluate the ability of the SIS model to identify high-risk cells, we compare its ranking performance against the PageRank baseline using the normalized discounted cumulative gain (nDCG) metric. As shown in Figure 11, the SIS model with the optimal τ consistently outperforms PageRank in most cases. The improvement is especially clear in the top 300 and top 500 ranked cells. This indicates that SIS is more effective at prioritizing nodes with high risk, which are often the most critical for fire prevention. One possible reason is that the SIS model incorporates contagion dynamics, where frequent influence from burning neighbors increase the steady-state risk of high risk areas. The optimal τ values in our results (around 16–50) are much higher than the epidemic threshold, which means the spreading is strong. In this situation, high risk areas are pushed to very high steady-state probabilities. This helps SIS bring the true high-risk cells to the top of the ranking. In contrast, weighted PageRank relies on linear aggregation of edge weights, making it less effective at distinguishing the high risk cells from less active ones.

As we extend the ranking window into the long-tail region, the performance gap between SIS and PageRank becomes smaller. For example, in SCC 2, the top 1000 ranks can cover a significant portion of its nodes (e.g., 60–80 percent). While these nodes remain part of a strongly connected component, they typically exhibit similarly low fire activity and contribute less significantly to the

5 RESULTS AND ANALYSIS

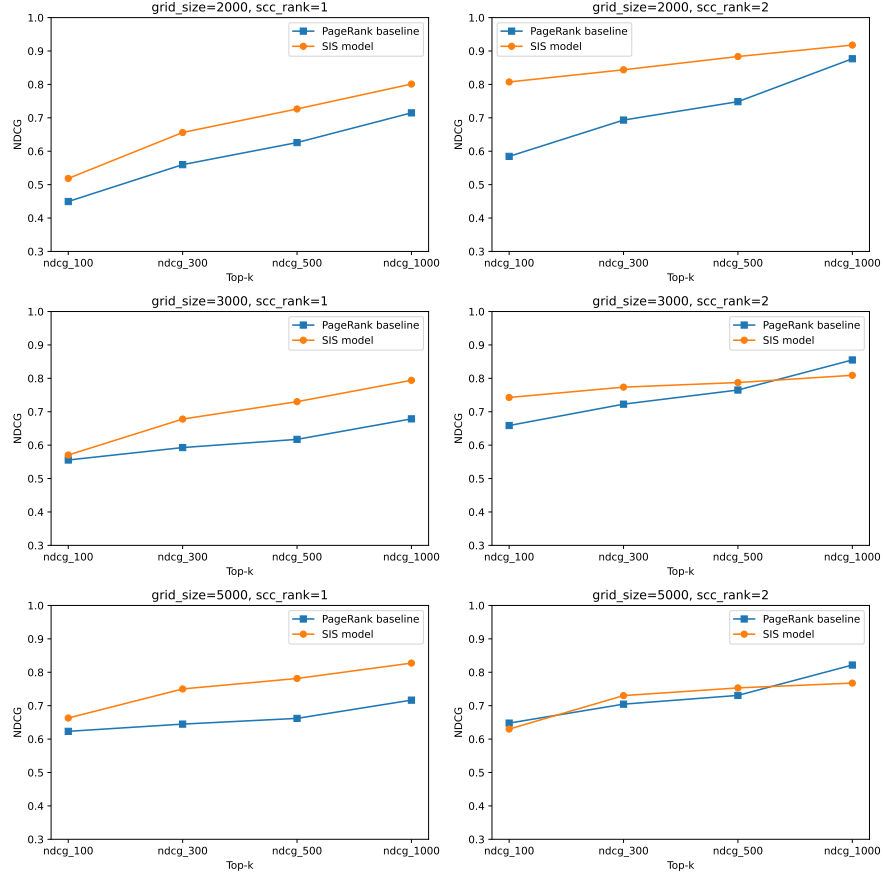


Figure 11: nDCG Performance of SIS and PageRank at Different Top-k Thresholds

core spread dynamics. These “tail nodes” thus form a functionally homogeneous structure where meaningful distinctions in risk are difficult to establish. Under such conditions, both SIS and PageRank tend to assign nearly uniform and low scores to these nodes, making it difficult to distinguish high-priority targets. While SIS’s dynamics amplify risk through repeated exposure, which are more meaningful in the densely connected core, these advantages become less relevant in the tail, where fire events are sparse and irregular. As a result, the advantage of SIS becomes smaller in these long-tail regions.

Specifically, at the 5000m grid size, we observe that in SCC 2, the SIS model shows a reduced advantage in nDCG compared to PageRank, even though it maintains a much higher Average Recall (AR) than in other cases. This performance gap highlights a trade-off between recall and ranking precision. The high AR in SCC 2 suggests that the SIS model is good at identifying a broad set of high-risk cells, especially those prone to repeated fire events. However,

its lower nDCG advantage implies that the model is less precise in ranking the top-risk cells. One possible explanation is that with a coarse grid, multiple fire events may be grouped into the same or neighboring cells, which blurs the distinction between the most critical and second-most critical nodes. In this case, the dynamic process in the SIS model may smooth out fine-grained differences between top nodes, while PageRank, relying on structural centrality, can provide a more accurate ranking of key nodes. This shows the importance of using multiple evaluation metrics and considering how spatial resolution and network structure together affect model's ability to differentiate fire risks.

5.3.3 Spatial Distribution of Fire Risk

Figure 12 and Figure 13 show the spatial distribution of actual fire risk v_{net} (left) and the steady-state fire risk assessed by the SIS model (right) with the optimal τ for grid size = 5000 m, in SCC 1 and SCC 2 respectively. These plots are used to qualitatively compare whether the SIS model can capture the spatial fire patterns at this resolution. We only show the results for grid size = 5000 m here because smaller grid sizes (e.g., 2000 m or 3000 m) contain too many cells, making the visualizations cluttered and harder to interpret. The 5000 m resolution provides a better visual balance between spatial detail and clarity.

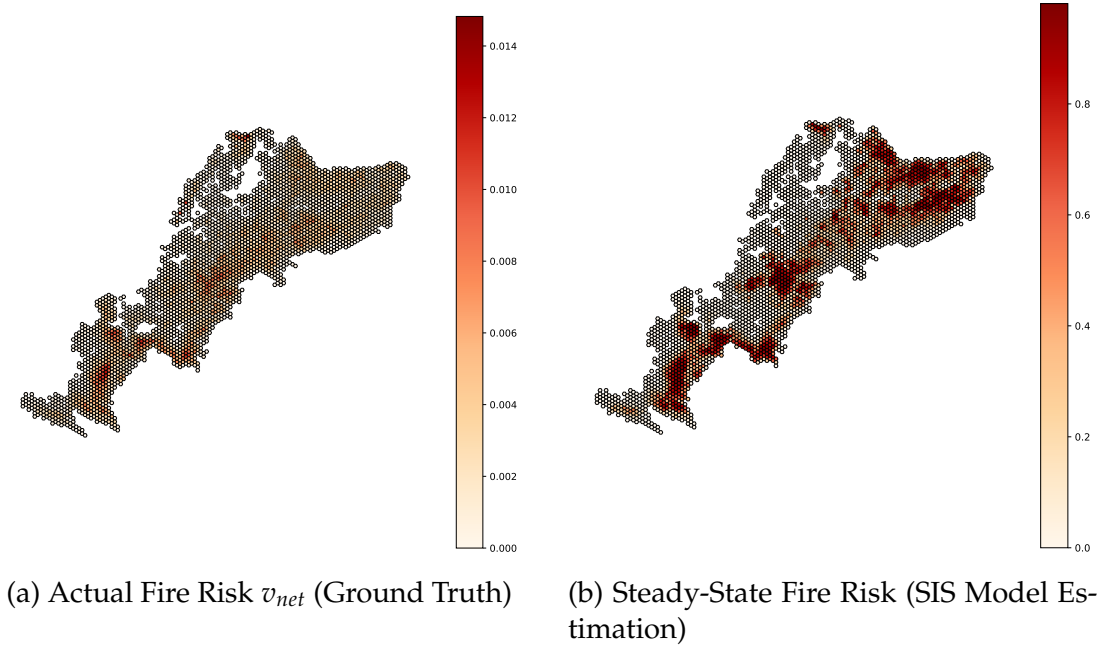


Figure 12: Grid Size = 5000, SCC 1: Spatial Distribution Comparison between Actual Fire Risk and SIS Model Risk Assessment

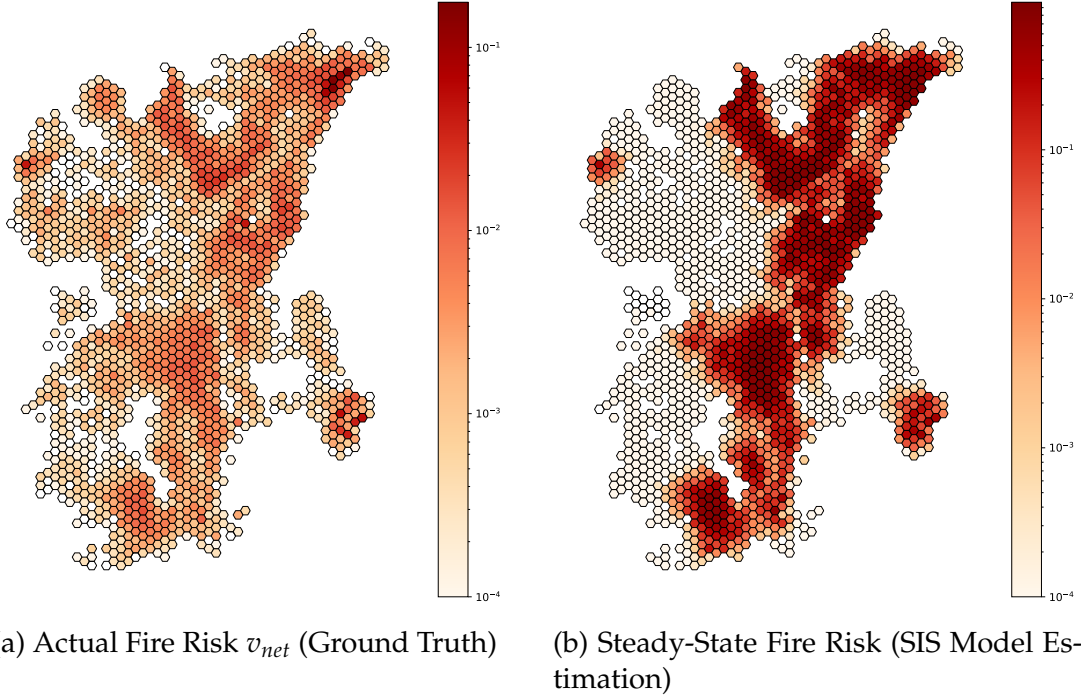


Figure 13: Grid Size = 5000, SCC 2: Spatial Distribution Comparison between Actual Fire Risk and SIS Risk Assessment

For both SCC 1 and SCC 2 at grid size=5000m, we observe a strong visual alignment between the high-activity regions in the actual fire distribution and the high-risk areas estimated by the SIS model. The model successfully identifies key fire hotspots, concentrating assessed risk in regions that historically experience frequent network-driven fire events. This spatial correspondence reinforces the SIS model’s capability to effectively rank and prioritize cells based on their long-term susceptibility to fire spread within the constructed Chronnet.

A notable difference lies in the scale and intensity of the values represented by the colorbars. The actual fire distribution displays relatively low probabilities, reflecting the sparse nature of fire occurrences over time. In contrast, the SIS model’s steady-state risk values are much higher, indicating a normalized or relative risk measure rather than a direct empirical probability. This distinction is consistent with our earlier findings regarding the trade-off between minimizing Jensen-Shannon Divergence (JSD) for distribution similarity and maximizing Average Recall (AR) for ranking performance. The SIS model tends to polarize risk scores, amplifying the distinction between highly active and less active nodes, which is beneficial for ranking purposes, even if it

does not precisely replicate the empirical probability distribution. Future work could focus on improving the model, perhaps by incorporating heterogeneous infection and recovery rates, to produce more accurate probabilistic forecasts without sacrificing its strong ranking performance.

Furthermore, the estimated steady-state distribution often appears more continuous and stronger in high-risk zones compared to the actual distribution. This is partly due to the relatively large τ values used which push many cells toward higher steady-state probabilities. As a result, the SIS model highlights recurrent fire pathways more clearly, while the observed fire events are discrete and tend to form more fragmented spatial patterns.

Comparing the two plots (SCC 1 vs. SCC 2) at grid size=5000m, we observe that SCC 2, although smaller in size, often exhibits a more spatially coherent high-risk region in the SIS estimation. In SCC 2, the steady state infection rate tend to cluster within a contiguous area rather than being scattered across the network. This spatial clarity likely contributes to its higher Average Recall performance, indicating that at this coarser resolution, SCC 2 captures a more structured and less noisy subset of fire spread patterns that the SIS model is better able to represent.

Overall, while some discrepancies exist, such as minor over or under estimation in certain localized areas, the overall alignment between estimated and observed patterns suggests that the SIS model provides a valuable tool for identifying high-priority regions for fire prevention and management efforts.

5.3.4 Impact of Self-Loops on SIS Model Performance

Section 5.1.3 showed that there is a high spearman correlation between the network-driven fire probability calculated with and without self-loops. The high correlation suggests that fire persistence and inter-cell spread are closely related phenomena. This raises the question of whether self-loops should be included in the SIS model's network structure. Since the model is designed to capture propagation, adding self-loops, which represent persistence over time rather than spread to neighboring nodes, might bias the estimation. To address this, we conduct a ablation study to assess the impact of self-loop edges on the SIS model's performance.

We created a modified version of the Chronnet graph by removing all self-loop edges. On these "no-self-loop" networks, we ran the same SIS model calibration and evaluation procedure as described in Section 4.3.2. We then compared the steady-state risk estimation from this modified model against the original ground truth.

The result of ablation study, presented in Table 7, show the robustness of the SIS modeling framework. The main finding is that removing self-loops

5 RESULTS AND ANALYSIS

Grid Size	SCC Rank	Model	Optimal τ	AR	JSD	nDCG@100	nDCG@300	nDCG@500	nDCG@1000
2000	1	Original SIS	16	0.67	0.62	0.52	0.66	0.73	0.80
		No-Self-Loop SIS	18	0.66	0.65	0.49	0.63	0.71	0.79
		PageRank	–	0.65	0.56	0.45	0.56	0.63	0.71
	2	Original SIS	25	0.75	0.53	0.81	0.84	0.88	0.92
		No-Self-Loop SIS	29	0.74	0.54	0.78	0.81	0.86	0.91
		PageRank	–	0.67	0.49	0.58	0.69	0.75	0.88
3000	1	Original SIS	48	0.74	0.38	0.57	0.68	0.73	0.79
		No-Self-Loop SIS	46	0.74	0.37	0.57	0.67	0.73	0.79
		PageRank	–	0.68	0.19	0.56	0.59	0.62	0.68
	2	Original SIS	17	0.77	0.16	0.74	0.77	0.79	0.81
		No-Self-Loop SIS	13	0.77	0.20	0.53	0.60	0.63	0.66
		PageRank	–	0.71	0.35	0.66	0.72	0.77	0.86
5000	1	Original SIS	29	0.76	0.22	0.66	0.75	0.78	0.83
		No-Self-Loop SIS	27	0.77	0.23	0.69	0.75	0.79	0.83
		PageRank	–	0.67	0.04	0.62	0.64	0.66	0.72
	2	Original SIS	23	0.82	0.18	0.63	0.73	0.75	0.77
		No-Self-Loop SIS	27	0.81	0.21	0.66	0.75	0.77	0.78
		PageRank	–	0.69	0.32	0.65	0.70	0.73	0.82

Table 7: Ablation study on the impact of self-loops.

edges from the Chronnet graph has a negligible impact on the model’s overall ranking performance. A comparison between the “Original SIS” and “No-Self-Loop SIS” models shows that the Average Recall (AR) scores remain very stable across all grid sizes and SCCs, with differences less than 0.01. This indicates that the model’s overall ranking capability is not dependent on information about fire persistence (self-loops) .

In addition, the performance in ranking the most critical high-risk nodes, as measured by nDCG, is mostly stable. At the 2,000m and 5,000m resolutions, the model’s performance changes only slightly when self-loops are removed. However, an unusual case occurs at the 3000m resolution for the second-largest SCC (SCC 2). In this case, the nDCG performance of the No-Self-Loop SIS model drops sharply. This outcome likely reflects a structural sensitivity: in this moderately connected SCC, self-loops provide the additional signal needed to distinguish top nodes.

More importantly, both SIS model outperform the PageRank baseline in most cases, particularly on the ranking metrics. This shows that the advantage of the SIS model is not due to self-loops but to its ability to capture the inter-cell spread. Overall, this ablation study validates our modeling approach. While the initial decision to include self-loops is justified because they represent a real-world fire persistence phenomenon, this study confirms that the model’s primary strength is its robustness. The main source of the model’s performance is its ability to simulate fire spread across the network, which remains strong

whether or not intra-cell persistence is included.

6 Conclusion

In this thesis, we developed and evaluated a data-driven framework to model fire spread by integrating *Chronnet*, a spatiotemporal event network, with the Susceptible–Infected–Susceptible (SIS) contagion model. We validated this framework in a case study of Colombia using VIIRS active fire data, and this method can be applied to any region with similar satellite observations, such as from MODIS or VIIRS. The goal was to explore whether concepts from epidemic modeling could offer a useful tool to understand fire spread patterns.

The main contribution is the Chronnet–SIS modeling framework. We transformed VIIRS 375 m active fire detections into a directed, weighted graph by binning events into 12-hour intervals and connecting fires that occurred in adjacent hexagonal grid cells. On this network, we applied the continuous-time N-Intertwined Mean-Field Approximation of the SIS model to estimate steady-state burning probabilities, treating fire spread as a recurrent contagion process similar to disease transmission.

To reduce the impact of satellite overpass bias and address the sparsity problem of chronnet graph, we aggregated fire events into two half-day bins. This improved the temporal consistency of the network while preserving the sequence of events. We also applied a pruning threshold to remove low-weight edges, which helped filter out noise and retain only plausible spread patterns.

To study how grid resolution influences the network, we tested a range of hexagonal sizes from 1 km to 7 km. The analysis showed that grids finer than 2 km led to fragmented, weakly connected structures, while coarser grids above 5 km tended to merge unrelated ignitions, producing misleading connections. The 2–5 km range appears to offer the best trade-off between preserving meaningful spatial relationships and minimizing noise. This result builds a practical guideline for future applications, suggesting an effective range to search for the most informative configuration when applying the framework to other region. In addition, at 2 km, 3 km, and 5 km resolutions, we observed clear small-world characteristics and heavy-tailed degree and strength distributions, suggesting the existence of hub-like areas with high fire recurrence.

One important observation is that only about 20–30% of fire events within the largest strongly connected components were actually connected in the network across all tested grid sizes. This suggests that most of the detected thermal anomalies in Colombia represent independent events rather than spatial spread, at least under the tested grid size range. Although the minimum grid size was constrained by the dataset resolution, using finer grids led to highly

fragmented networks with very low connectivity, as indicated by the low ratio of nodes in the largest strongly connected component. Such sparsity did not support meaningful propagation modeling. The SIS model should therefore be understood as describing the dynamics of spread where connections are detectable, rather than modeling all ignition events. This validates our method of distinguishing between spontaneous and network-driven fire during evaluation.

For the SIS modeling, we calibrated the effective infection rate τ of the SIS model by minimizing the Jensen–Shannon divergence between estimated and observed distributions, and maximizing average recall for top-ranked cells. While PageRank served as a baseline for static centrality, the SIS model consistently achieved higher Average Recall (AR), suggesting it has a superior ability to rank influential nodes dynamically. The improvement is particularly evident in the top-ranked cells, which are mostly critical for prevention efforts. These results demonstrate that fire spread risk is better captured through dynamic propagation modeling than by static centrality measures. We further confirmed the model’s robustness and demonstrated that its advantage derives from capturing the complex contagion dynamics of inter-cell spread, rather than simply relying on the strong signal of fire persistence (self-loops).

This study also has several limitations. First, while the model is good at ranking which areas are at high risk, its estimated probability values are not numerically aligned with the real-world data. This partly because we used homogeneous parameters for infection and recovery rates, which does not reflect spatial differences in vegetation, climate, or fire management capacity. Second, to address the sparsity problem in chronnet construction, we select the fixed 12-hour temporal bins for satellite compatibility, but this could obscure finer-scale dynamics, especially in fast-moving fires. Third, our model did not incorporate environmental variables known to influence fire behavior, such as wind speed and direction, elevation, and fuel availability, which limits the ecological realism of the model. Fourth, dividing fire events into just two categories, that are “spontaneous” and “network-driven”, could oversimplify the spread mechanism. In this study, a fire is labeled as network-driven if a neighboring cell was burning in the previous 12-hour window; otherwise, it is considered spontaneous. This definition assumes that temporal succession and spatial proximity are enough to imply a causal link. However, it fails to distinguish true propagation from events where a common external factor (e.g., lightning, arson) ignites multiple nearby cells independently. More importantly, this framework treats spontaneous fires as events outside the model’s core dynamic, when in reality, they are the very ignition points that initiate the network-driven cascades. By learning only from established propagation events, the model’s ability to forecast the emergence of new fire clusters is inherently limited.

Future work could address these gaps. One direction is to integrate more data, such as burned area data. We can use this to group scattered fire hotspots into single, complete fire events. This would create a network that better reflects real-world fire spread and would help solve the problem of network sparsity. With a more reliable network, the next step is to make the model's rules more flexible. We should allow the rates of fire spread and extinction (the parameters β and γ) to vary based on local conditions, such as vegetation type or weather, instead of using the same values everywhere. Additionally, we need to better simulate how fires start. Future models should not only focus on how fires spread but should also include a component that assess the risk of "spontaneous" ignitions. A simple way to do this is to add a term representing ignition risk directly into the existing SIS model equations. Finally, to enhance the model's practical value, future work should also explore combining it with physical models, using it for real-time forecasting, and quantifying the uncertainty in its estimation to better support real-world decision-making.

Overall, this thesis translates the dynamics of fire spread into a network-based contagion framework. Although simplifications were necessary, especially regarding environmental complexity, the results show that Chronnet-SIS modeling framework can offer valuable insights into how fires spread and how their risks might be quantified. This approach could serve as a foundation for developing more adaptive, data-driven tools to support fire risk management at large scales.

Appendix

6.1 Additional Figures



Figure 14: Node Centrality vs. Network-Driven Fire Probability

6.2 Code Availability

The source code for this study is available at: https://github.com/mabelhu465/2025_chronnet_sis/tree/main

References

- [1] Kendra K McLauchlan, Philip E Higuera, Jessica Miesel, Brendan M Rogers, Jennifer Schweitzer, Jacquelyn K Shuman, Alan J Tepley, J Morgan Varner, Thomas T Veblen, Solny A Adalsteinsson, et al. Fire as a fundamental ecological process: Research advances and frontiers. *Journal of Ecology*, 108(5):2047–2069, 2020.
- [2] Andrew L Sullivan. Wildland surface fire spread modelling, 1990–2007. 1: Physical and quasi-physical models. *International Journal of Wildland Fire*, 18(4):349–368, 2009.
- [3] Andrew L Sullivan. Wildland surface fire spread modelling, 1990–2007. 2: Empirical and quasi-empirical models. *International Journal of Wildland Fire*, 18(4):369–386, 2009.
- [4] Andrew L Sullivan. Wildland surface fire spread modelling, 1990–2007. 3: Simulation and mathematical analogue models. *International Journal of Wildland Fire*, 18(4):387–403, 2009.
- [5] Wenyu Jiang, Fei Wang, Guofeng Su, Xin Li, Guanning Wang, Xinxin Zheng, Ting Wang, and Qingxiang Meng. Modeling wildfire spread with an irregular graph network. *Fire*, 5(6):185, 2022.
- [6] Moritz Rösch, Michael Nolde, Tobias Ullmann, and Torsten Riedlinger. Data-driven wildfire spread modeling of european wildfires using a spatiotemporal graph neural network. *Fire*, 7(6):207, 2024.
- [7] Vera LJ Somers and Ian R Manchester. Priority maps for surveillance and intervention of wildfires and other spreading processes. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 739–745. IEEE, 2019.
- [8] Joan Boters Pitarch, María Teresa Signes Pont, Julian Szymański, and Higinio Mora Mora. Application of a stochastic compartmental model to approach the spread of environmental events with climatic bias. *Ecological Informatics*, 77:102266, 2023.
- [9] Leonardo N Ferreira, Didier A Vega-Oliveros, Moshé Cotacallapa, Manoel F Cardoso, Marcos G Quiles, Liang Zhao, and Elbert EN Macau. Spatiotemporal data analysis with chronological networks. *Nature communications*, 11(1):4036, 2020.

REFERENCES

- [10] Philip E Paré, Carolyn L Beck, and Tamer Başar. Modeling, estimation, and analysis of epidemics over networks: An overview. *Annual Reviews in Control*, 50:345–360, 2020.
- [11] N Hoyos, A Correa-Metrio, A Sisa, MA Ramos-Fabiel, JM Espinosa, JC Restrepo, and J Escobar. The environmental envelope of fires in the colombian caribbean. *Applied geography*, 84:42–54, 2017.
- [12] Sofia Bajocco, Nikos Koutsias, and Carlo Ricotta. Linking fire ignitions hotspots and fuel phenology: The importance of being seasonal. *Ecological Indicators*, 82:433–440, 2017.
- [13] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford infolab, 1999.
- [14] Jingjing Sun, Wenwen Qi, Yuandong Huang, Chong Xu, and Wentao Yang. Facing the wildfire spread risk challenge: where are we now and where are we going? *Fire*, 6(6):228, 2023.
- [15] Harikesh Singh, Li-Minn Ang, Tom Lewis, Dipak Paudyal, Mauricio Acuna, Prashant Kumar Srivastava, and Sanjeev Kumar Srivastava. Trending and emerging prospects of physics-based and ml-based wildfire spread models: A comprehensive review. *Journal of Forestry Research*, 35(1):135, 2024.
- [16] Mark A Finney. Fire growth using minimum travel time methods. *Canadian Journal of Forest Research*, 32(8):1420–1424, 2002.
- [17] L Hernández Encinas, S Hoya White, A Martín Del Rey, and G Rodríguez Sánchez. Modelling forest fire spread using hexagonal cellular automata. *Applied mathematical modelling*, 31(6):1213–1227, 2007.
- [18] Paul Johnston, Joel Kelso, and George J Milne. Efficient simulation of wildfire spread on an irregular grid. *International Journal of Wildland Fire*, 17(5):614–627, 2008.
- [19] Alexander Stepanov and James MacGregor Smith. Modeling wildfire propagation with delaunay triangulation and shortest path algorithms. *European Journal of Operational Research*, 218(3):775–788, 2012.
- [20] Mohammad Hajian, Emanuel Melachrinoudis, and Peter Kubat. Modeling wildfire propagation with the stochastic shortest path: A fast simulation approach. *Environmental modelling & software*, 82:73–88, 2016.

REFERENCES

- [21] Sanjay Kumar, Muskan Saini, Muskan Goel, and Nipun Aggarwal. Modeling information diffusion in online social networks using sei epidemic model. *Procedia Computer Science*, 171:672–678, 2020.
- [22] Fabio Caccioli, Paolo Barucca, and Teruyoshi Kobayashi. Network models of financial systemic risk: a review. *Journal of Computational Social Science*, 1(1):81–114, 2018.
- [23] Alberto Ceria, Klemens Köstler, Rommy Gobardhan, and Huijuan Wang. Modeling airport congestion contagion by heterogeneous sis epidemic spreading on airline networks. *Plos one*, 16(1):e0245043, 2021.
- [24] Maksim Kitsak, Lazaros K Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H Eugene Stanley, and Hernán A Makse. Identification of influential spreaders in complex networks. *Nature physics*, 6(11):888–893, 2010.
- [25] An Zeng and Cheng-Jun Zhang. Ranking spreaders by decomposing complex networks. *Physics letters A*, 377(14):1031–1035, 2013.
- [26] Duan-Bing Chen, Hui Gao, Linyuan Lü, and Tao Zhou. Identifying influential nodes in large-scale directed networks: the role of clustering. *PloS one*, 8(10):e77455, 2013.
- [27] Linyuan Lü, Yi-Cheng Zhang, Chi Ho Yeung, and Tao Zhou. Leaders in social networks, the delicious case. *PloS one*, 6(6):e21202, 2011.
- [28] Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446, 2002.
- [29] Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, 37(1):145–151, 2002.
- [30] Piet Van Mieghem, Jasmina Omic, and Robert Kooij. Virus spread in networks. *IEEE/ACM Transactions On Networking*, 17(1):1–14, 2008.
- [31] Saba Akram and Quarrat Ul Ann. Newton raphson method. *International Journal of Scientific & Engineering Research*, 6(7):1748–1752, 2015.
- [32] Luis A Nunes Amaral, Antonio Scala, Marc Barthelemy, and H Eugene Stanley. Classes of small-world networks. *Proceedings of the national academy of sciences*, 97(21):11149–11152, 2000.
- [33] Márton Pósfai and Albert-László Barabási. *Network science*, volume 3. Cite-seer, 2016.