Forecasting landslides using community detection on geophysical satellite data

Vrinda D. Desai, Farnaz Fazelpour, Alexander L. Handwerger, and Karen E. Daniels

Forecasting landslides using community detection on geophysical satellite data

Vrinda D. Desai and Farnaz Fazelpour
Physics Department, North Carolina State University, Raleigh, NC, USA

Alexander L. Handwerger
Joint Institute for Regional Earth System Science and Engineering,
University of California Los Angeles, Los Angeles, CA, USA and Jet Propulsion Laboratory,
California Institute of Technology, Pasadena, CA, USA

Karen E. Daniels
Physics Department, North Carolina State University, Raleigh, NC, USA
(Dated: May 16, 2023)

As a result of extreme weather conditions, such as heavy precipitation, natural hillslopes can fail dramatically; these slope failures can occur on a dry day due to time lags between rainfall and pore-water pressure change at depth, or even after days to years of slow-motion. While the pre-failure deformation is sometimes apparent in retrospect, it remains challenging to predict the sudden transition from gradual deformation (creep) to runaway failure. We use a network science method – multilayer modularity optimization – to investigate the spatiotemporal patterns of deformation in a region near the 2017 Mud Creek, California landslide. We transform satellite radar data from the study site into a spatially-embedded network in which the nodes are patches of ground and the edges connect the nearest neighbors, with a series of layers representing consecutive transits of the satellite. Each edge is weighted by the product of the local slope (susceptibility to failure) measured from a digital elevation model and ground surface deformation (current rheological state) from interferometric synthetic aperture radar (InSAR). We use multilayer modularity optimization to identify strongly-connected clusters of nodes (communities) and are able to identify both the location of Mud Creek and nearby creeping landslides which have not yet failed. We develop a metric, community persistence, to quantify patterns of ground deformation leading up to failure, and find that this metric increases from a baseline value in the weeks leading up to Mud Creek's failure. These methods promise as a technique for highlighting regions at risk of catastrophic failure.

I. INTRODUCTION

Slow-moving landslides creep downhill at rates from millimeters to several meters a year [1], a process that sometimes proceeds for hundreds of years [2, 3]. Yet, a slow-moving landslide can reach runaway acceleration, and in many cases this is likely to be triggered by heavy precipitation, as happened in the 2017 Mud Creek landslide located on the Big Sur Coast, CA. The processes which influence these transitions are myriad, and additionally include topographic slope, hydrological and mechanical properties, and natural or man-made disturbances. Advancement in remote sensing techniques allows for the detection of ground deformation at spatiotemporal resolutions that were not achievable a decade ago [4, 5]. Previous work in landslide forecasting has focused on either (1) estimating the location and shape of a landslide using spatial analysis of slope stability [6] or (2) the time of failure using temporal analysis of ground deformation [7]. There has been recent progress on forecasting both location and time [5, 8]. Linking kinematic data and the underlying micro-mechanics of granular failure together has shown to be effective in forecasting the location and time of granular failure [5, 9, 10].

Interferometric processing of synthetic aperture radar (InSAR) makes studies of landslide and other ground deformation possible at high spatial and temporal resolu-

tion. Ground deformation over large surface areas of the Earth is measured with centimeter to millimeter accuracy using InSAR [11] by exploiting the phase difference between two SAR snapshots acquired over the same area at two different times. This quantifies the amount of ground deformation during that time interval, projected along the radar line-of-sight (LOS) direction [12, 13]. InSAR techniques have been applied in the study of landslides and are starting to be incorporated into early warning systems (EWS)[6, 13–16].

EWS have used these advancements in radar monitoring data to monitor areas of known risk and identify regions heavily affected by disaster [17], but there are still significant drawbacks and limitations to the forecasting of landslides and other hazards [9]. One such drawback is the use of thresholds that are defined to predict the onset of failure in landslides [18], and recent studies have shown that these thresholds are site-specific and subjective [9]. A common practice is to only consider the region that is accelerating rapidly [18], but important information about the interactions occurring between the region and the surrounding area then go undetected [5, 19, 20].

The transition of a creeping landslide to a catastrophic failure is complex and the system's behavior emerges from dynamics that are occurring beneath the surface [21]. The landslide system is modeled as a collection of patches that are interacting with each other in such a way that the emerging behaviour is not immediately pre-

dictable [22], which is why it is crucial to not only look at the immediate region, but also the surrounding area. Thus, one of the biggest challenges to EWS in regards to landslides is harnessing the full potential of these data sets using the analysis and interpretation of the underlying micro-mechanics in the time before granular failure [5].

Recent studies have shown that granular failure does not occur spontaneously, but instead there is a transitional period between stable deformation and catastrophic failure, which is indicated by a distinct dynamical pattern, influenced by kinematics[8, 9, 23]. Consider a region that consists of multiple closely connected patches, or clusters. As the region approaches failure, the spatiotemporal dynamics have been characterized into two pieces: (1) similarity within a cluster increases as they begin to move together and (2) similarity between clusters decreases [8]. The transition to imminent failure is guided by the spatiotemporal persistence of community C, followed by a sharp increase in the motion, or velocity, of C in respect to the rest of the system, or area [8].

Our goal is to use kinematic data retrieved from remote sensing techniques to find patches of ground, defined as communities, that are likely to fail and their impending time of failure. To develop a holistic approach that integrates the dynamics of ground deformation, along with the physics of slope stability, we use network science techniques of micro-scale granular failure (grains on the order of 1 mm to 1 cm) [24–26] and apply it on a macroscale system (patches of area on the order of 7x7 to 20x20 m) [8]. A network framework gives us the ability to infer the structure and underlying dynamics of a creeping landslide using information such as displacement that can be retrieved using remote sensing techniques [22].

A. Study Site: Mud Creek vicinity

Within the California Coast Ranges, landslides are abundant due to mechanically weak rocks, active uplift, and high seasonal precipitation [27–30]. More than 650 active slow-moving landslides have been identified and mapped within the region, but few of them have failed catastrophically [29, 30]. On a dry day (May 20, 2017) following a long period of heavy rainfall that ended on April 17, 2017, the Mud Creek landslide, which had been slowly moving for decades, failed catastrophically and destroyed part of California State Highway 1 (CA-1) [27, 31].

The Central Coast Ranges, which contain Mud Creek, receive about ~ 1 m of rain per year, of which 80% falls between October and May [27, 33]. The precipitation in this region varies significantly from year to year due to changes in the frequency and strength of atmospheric rivers; climate models predict that California will experience a dramatic increase in frequency of shifts from extreme dry seasons to extreme wet seasons over the next couple of decades [33]. Since landslide dynamics can be
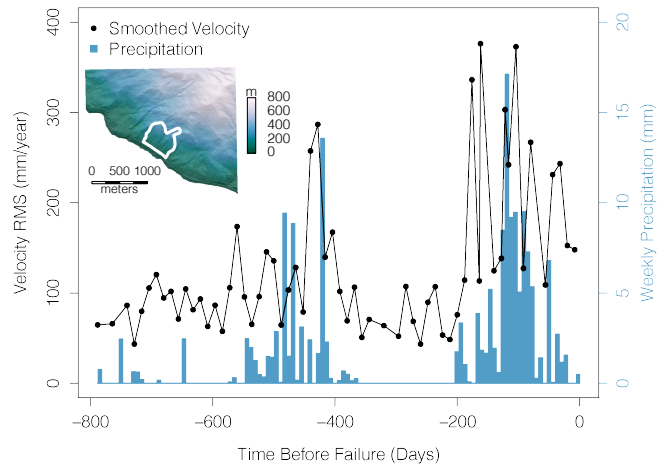
FIG. 1. The root mean square (RMS) of the line of sight (LOS) velocity (black circles) for the entire study area leading up to the day of failure ($T = 0$); data from Handwerger et al. [27]. Blue bars are the total weekly precipitation, from PRISM [32]. The wet seasons correspond to roughly December to May each year.

driven by seasonal changes, as seen in Fig. 1, the predicted increase in changes from dry to wet seasons is a cause for concern [27, 33, 34].

Changes in the shear strength of a hillslope are dependent on the water-saturation level of the material [35–37]. If the material is not fully saturated, then capillary effects increase shear strength and help stabilize a hillslope [35, 36, 38]. On the other hand, if the material is almost or fully saturated, then an increase in the pore-water pressure leads to a reduction in the shear strength of the material [36, 37], and may possibly lead to instability within the hillslope [21, 35, 37, 38]. Note that the pore-water pressure varies strongly with depth, as rain permeates the hillslope over time. In addition, the parameters that influence slope stability originate from critical-state soil mechanics with a Mohr-Coulomb failure criterion, which assumes that failure is controlled by maximum shear strength [39, 40]. Saturated material, corresponding to high pore-water pressure, plus steep slopes increases the susceptibility to slope failure [21].

In this paper, we examine the area in and around Mud Creek (3 km × 3 km), shown in Fig. 1, to explore the dynamics within Mud Creek and compare it to the rest of the area, which consists of 2 other creeping landslides that did not fail. We are interested in analyzing, and with the eventual aim of forecasting, the transition of a slow-moving landslide to sudden catastrophic failure.

B. Comparing Mud Creek to Other Regions

Since there is only one creeping landslide that catastrophically failed within the study area, we create a set of controls to test for sensitivity and specificity. For this,
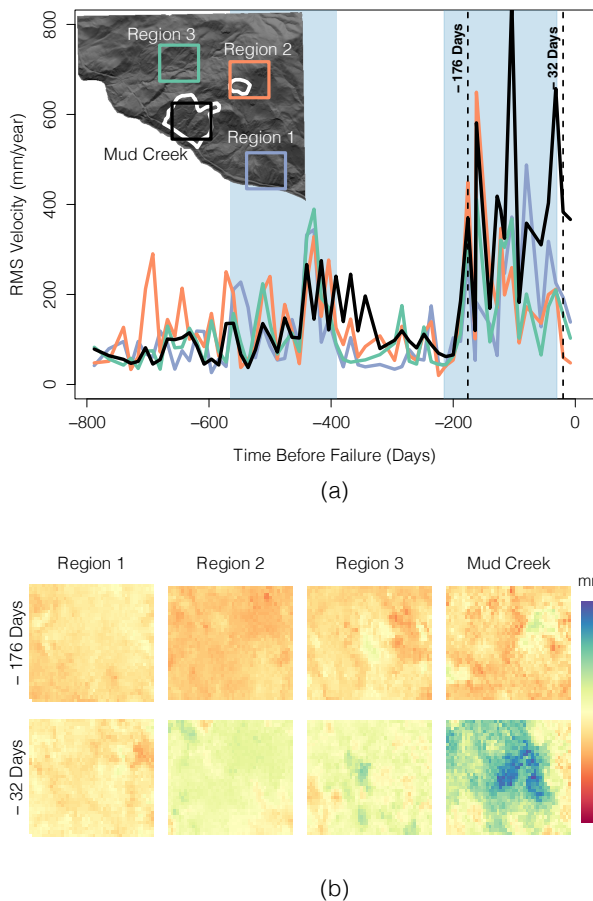
FIG. 2. (a) The root mean square (RMS) of the line of sight (LOS) velocity, for each of the 4 regions. The wet seasons shown in Fig. 1 are shaded in blue. Inset: bare earth hillshade map from lidar showing the four regions within the study area. (b) Maps of the LOS velocity for each of the 4 regions at two times: -32 and -176 days before failure. Data is from [27].

we additionally selected three regions within the study area to analyze alongside Mud Creek (see Fig. 2). Of these, Region 2 is a creeping landslide that was not identified in Handwerger *et al.* [27], but instead was recently observed with airborne InSAR data (see Fig. 2); and Regions 1 and 3 are selected because they are steep areas that are not creeping. For each of these regions, the LOS velocity (from Sentinel-1 InSAR data, described in more detail in §II B), plotted in Fig. 2, shows that there is not a significant difference between them and Mud Creek for the majority of the time period. In Fig. 2(b), the velocities for the four regions are mapped at two different time steps (-176 and -32 days) to show that the regions are moving similarly to each other at -176 days and later at -32 days Mud Creek is moving significantly faster than the other three regions.

C. Networks

It is difficult to quantify the structural changes happening beneath the surface, and thereby forecasting when and where a landslide is likely to occur. Thus, it is desirable to turn to reduced descriptions of the system. One such technique, network science, has been successful at quantifying and understanding the spatiotemporal dynamics of complex systems without the inclusion of a detailed model of the internal interactions. Previous studies, such as [41, 42], have shown that a network clustering technique known as *community detection* (reviewed in Fortunato [43]) can be used to extract hidden relations among nodes in static single-layer networks. Recent studies, such as the analysis of the evolution of the force networks within granular materials by Papadopoulos *et al.* [44], have extended that technique to be used in *multilayer networks* to analyze spatially-embedded networks (*i.e.* networks in which nodes only share edges with their own neighbors) that are more similar to earth materials.

Network science provides a way to characterize this complex system for consistent patterns of causality via community-detection techniques. For each snapshot of ground deformation, there is a correlating static network layer that is connected together into a multilayer network, where a simplistic version of a multilayer network is shown in Fig. 3. We analyze the transition of a slow-moving landslide, as represented in a multilayer network, to sudden catastrophic failure using *community persistence*, a measure we have developed to quantify the consistency of community structures. Using this measure, we are able to narrow down the period in which Mud Creek could be said to be at risk of failure, compared to what would be possible using InSAR alone.

II. DATA ASSIMILATION

We use two types of data to create our multilayer network: digital elevation models (DEM) for spatial information and slope and displacement snapshots from InSAR to estimate the current rheological state. This section details the process by which we define our multilayer network from the DEM and InSAR data; the edges within each time step (layer ℓ) are weighted by both the local slope (steeper slopes are more susceptible to failure) and the local creep velocity (faster speeds indicate a current soil state with more fluid rheology). Local slope is held fixed over the duration of the analysis.

A. Digital Elevation Model

We measure the elevation at all positions in the study site using a Digital Elevation Model (DEM) at 1 m resolution, downloaded from OpenTopography (May 17 2021), chosen because it penetrates vegetation and measures the
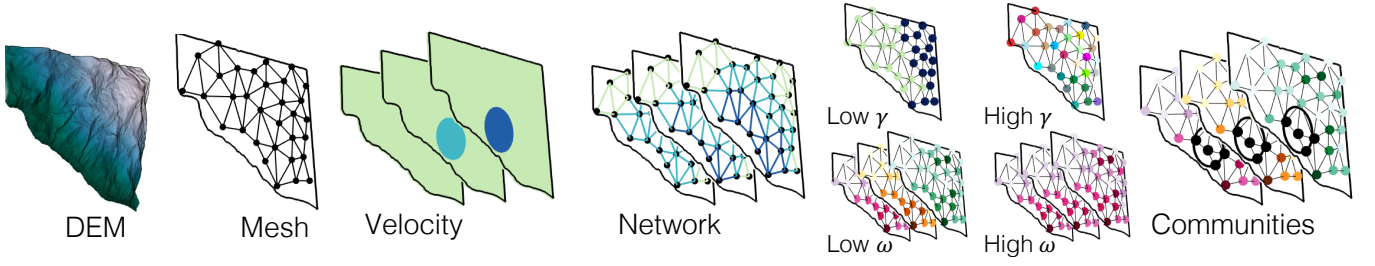
FIG. 3. Schematic of the construction and analysis of our multilayer network. We sample the digital elevation model (DEM) and a series of velocity maps on a disordered mesh to create a multilayer network weighted by both. Using this network, we selected the spatial coupling parameter γ and the temporal coupling parameter ω to detect communities of interest.

true ground surface. This DEM is from the USGS, 3D Elevation Program: ARRA-CA CentralCoast-Z3 2010 [45]. We define the elevation as a field $h(\vec{r})$ with position $\vec{r} = (x, y)$ given in UTM coordinates. From this, we calculate the slope edge weight as $s = \frac{|\Delta h|}{|\Delta \vec{r}|}$, which has a flat distribution for $0° \leq s \leq 18°$, after which steeper slopes are rarer, up to a maximum of $32°$ (such as within Mud Creek). The mean slope weight for the other three regions are: $25°$, $24°$, $25°$. Note that the study site contains steeper slopes, but those are averaged out when calculating the edge weights.

B. InSAR

We utilize an existing dataset from Handwerger *et al.* [27], comprising InSAR snapshots from the Copernicus Sentinel-1 A/B satellites (see Handwerger *et al.* [27] for InSAR processing details). The resulting series of snapshots provides a total of 63 layers with 10 m $\times$ 12 m resolution estimating ground deformation in a 3 km $\times$ 3 km region in the vicinity of Mud Creek area from February 2015 to May 2017. Each InSAR snapshot provides LOS displacements, where the direction of motion is either towards or away from the satellite. Using the InSAR data, we have a layer ℓ for each snapshot within the multilayer network. These layers span 6, 12, or 24 days due to either missed orbits (repeat cycle is 6 days) or low coherence.

For each layer ℓ (corresponding to an InSAR snapshot), we defined displacement as a field $U(x, y, \ell)$. Cumulative displacement is the displacement of the surface from the first reference frame ($\ell = 0$), for which $U(x, y, 0) = 0$ everywhere. Subsequent layers contain the cumulative displacement that has happened during layers 0 to ℓ. To quantify deformation, we calculate the incremental displacement u between pairs of adjacent snapshots as $u(x, y, \ell) = U(x, y, \ell) - U(x, y, \ell - 1)$. Since the acquisition interval for the satellites varied from 6, 12, and 24 days, we converted incremental displacement into velocity, $v(x, y, \ell) = \frac{u(x,y,\ell) - u(x,y,\ell-1)}{\Delta t}$, where Δt is time between any 2 consecutive snapshots. The noise from the InSAR data (not visualized in Fig. 3) increases when we take the derivative of the LOS displacement; this obscures motion that is otherwise de-

tected when retroactively looking at cumulative displacement maps. In order to reduce noise within each layer of the original dataset, we apply a a 3×3 cell matrix of a Gaussian smoothing kernel.

C. Defining the multilayer network

Our analysis takes place on a multilayer network, where each layer ℓ represents a time in space (corresponding to an InSAR snapshot). Each layer is represented as a spatially embedded network consisting of nodes and edges. Each node represents a patch of area, and each edge is the connection between patches of areas that are neighbors. This network consists of the same nodes and edge connections for all layers within the system, which we define as a mesh.

To create the spatially embedded network, we sampled the DEM at $n \sim 5000$ nodes for the study area shown in Fig. 2, selected to be large enough to be computationally efficient but small enough to provide spatial resolution within ridges and valleys. We considered uniform, random, and uniform random grids to create a mesh that would not introduce artefacts into the results or lead to mesh-specific results. Using Poisson disc sampling [46], a type of uniform random grid, we randomly selected node locations spaced an average of 20 m apart, without any preferred orientation, as shown in Fig. 3. Pairs of adjacent nodes are connected by edges using Delaunay triangulation (selects the closest nodes to each node), defining our binary adjacency matrix B_{ij}. We created an ensemble of 10 such meshes that are uniformly sampled across the area while also allowing us to test for sensitivity of the results to the mesh that was used.

Information about the current state of rheology and the susceptibility of failure was incorporated into the network using data from the DEM and InSAR to weight the edges. For each edge ij, we calculated the slope $s_{ij} = \frac{|h_i - h_j|}{|\vec{r}_i - \vec{r}_j|}$, where h_i is the average elevation (given by the DEM) of the patch of area represented by node i and $\vec{r}_i$ is the central coordinates of the patch of area, or node i. For an InSAR snapshot (layer ℓ), each node i has some velocity $v_{i\ell}$ that represents the backward-difference velocity, selected since it is convenient for use

in real-time monitoring. We then calculated velocity for each edge ij as the absolute average velocity defined as $|\bar{v}_{ij\ell}| = |\frac{v_{i\ell} + v_{j\ell}}{2}|$. The edges are weighted as $s_{ij}|\bar{v}_{ij\ell}|$ for each layer ℓ, where s_{ij} is constant for all layers, but $|\bar{v}_{ij\ell}|$ changes for each layer ℓ. We define our adjacency matrix as

$$A_{ij\ell} = s_{ij}|\bar{v}_{ij\ell}|B_{ij} \tag{1}$$

for a single layer ℓ as a snapshot in time, with node indices i, j specifying spatial location.

Our computer code is available on GitHub [58].

III. NETWORK ANALYSIS

A *community* within a network is a collection of nodes that are more strongly related to each other than to other nodes (see Fig. 3) [43, 47, 48]. We seek to identify whether communities are present within the Mud Creek dataset, and evaluate whether there are particularly unstable regions or periods within the system. For example, clusters of nodes that are connected by edges, and moving at relatively higher speeds and/or on steeper slopes, could be identified as a community. The community structure of a multilayer network shows the evolution of the communities over time, where communities are formed from partitioning the graph based on a defining characteristic [47].

A. Community detection

In this study, we analyze our multilayer network (Eq. 1) using a method based on optimizing the modularity Q; the method partitions nodes into communities (each community is thereby given an arbitrary ID by the algorithm to identify it) where the edge weights within the community are stronger than one would expect in comparison to a null model[47, 49, 50]. We optimize the modularity of the community structures using the NetWiki GenLouvain implementation for multilayer networks [48, 51]. The multilayer modularity is calculated as

$$Q = \frac{1}{2\mu} \sum_{ij\ell m} [(A_{ij\ell} - \gamma P_{ij\ell})\delta_{\ell m} + \omega_{j\ell m}\delta_{ij}]\delta(c_{i\ell}, c_{jm}), \tag{2}$$

where μ is the sum of all edge weights in the network, $c_{i\ell}$ is the community of node i, $c_{j\ell}$ is the community of node j, γ is the *structural resolution parameter*, $\omega_{j\ell m}$ is the *interlayer coupling*, and $P_{ij\ell}$ is the expected edge weight between nodes i and j under a specified *null model* in layer ℓ [48]. Note that using a global null model disregards the temporal fluctuations and does not improve the results of the community detection.

To optimize the modularity, the nodes are partitioned into communities in such a way that the total connection strength — the cumulative weight of connected edges —

within a community is much greater than what would be expected from a null model. For any network, the choice of null model provides the expected weight of a randomly-selected pair of nodes, and affects the interpretation of the network [52]. Because our system is a spatially-embedded network, the work of Bassett *et al.* [44, 53] guides our choice of the *geographical null model* rather than the more common Newman-Girvan null model [49, 50] for which all edges are equally connected to each other. The geographical null model is defined as

$$P_{ij\ell} = \rho_\ell B_{ij}, \tag{3}$$

where $\rho_\ell = \langle s_{ij}|v_{ij\ell}|\rangle_{ij}$ is the mean edge weight of network layer ℓ (averaged over all ij edges) and B_{ij} is the binary adjacency matrix, which is constant through all layers and encodes the geophysical proximity.

There are two resolution parameters used to calculate modularity Q, both of which are selected to control the total number of communities detected, as seen in Fig. 3. The structural resolution parameter γ sets the weight of the geographic null model. Within each layer ℓ, γ sets the spatial resolution at which nodes are partitioned into communities, via comparison to the mean edge weight ρ_ℓ; higher values of γ will yield more communities that are smaller in size [44]. The corresponding temporal resolution parameter is contained within the interlayer coupling $\omega_{j\ell m}$, which is the weight of the edge coupling node j to itself in layers (ℓ, m). Here, we take $\omega_{j\ell m} = \omega$ (constant), for simplicity. We want to consider ω values that take into account the strain-loading history. Higher values of ω more strongly favor the continuity of communities across the series of layers, while for $\omega = 0$, each layer is considered independently (no communities span the layers), and thus no consideration of the strain-loading history.

B. Selecting the coupling parameters (γ, ω)

We wish to select resolution parameters that highlight communities within the network which (1) have stronger weights than average and (2) continue that behavior over some time interval. To satisfy the former, we pick $\gamma = 1$; this value favors communities with stronger-than-average weights (larger than ρ_ℓ) [44]. In contrast, the interlayer coupling ω is normally of order 1 [44]. For the latter, we want to find a value of ω which is large enough that important communities persist over time, but not so large that we fail to detect changes in the temporal state of the study area. Since there is not yet consensus in the literature on specific values of ω, we explore two measures which have been useful in earlier studies: *flexibility* and *stationarity* [41, 44, 54].

We quantify the local properties of the nodes using a measure known as flexibility [54]. It is the measure of the number of layers, g_i, that each node i changes community allegiance over the total possible layers ($L - 1$). The

flexibility of the multilayer network as a whole for all nodes N is defined as:

$$\Xi = \frac{1}{N} \sum_i \frac{g_i}{L-1}.$$ (4)

To capture the global properties of the communities found within the network, we measure the consistency with which nodes are placed in communities over time using a measure known as stationarity [41, 55]. As given in Papadopoulos *et al.* [44], the stationarity of a multilayer network is ζ for all nodes N:

$$\zeta = \frac{1}{C} \sum_c \frac{\sum_{\ell_i}^{\ell_f-1} J(c_\ell, c_{\ell+1})}{\ell_f - \ell_i - 1},$$ (5)

where C is the total number of communities and $J(c_\ell, c_{\ell+1})$ is the joint fraction (the ratio of the number of mutual nodes in community c in layers ℓ and $\ell+1$ to the number of nodes in the union of layers ℓ and $\ell+1$). The stationarity ζ approaches 1 at high values of ω. In our calculation, we exclude communities that never contain more than one node [44].

Fig. 4 shows the values of both the flexibility (Eq. 4) and the stationarity (Eq. 5), repeated for each of the 10 meshes. In order to be sensitive to both the intra- (spatial) and inter- (temporal) layer connections, ω should also be on the same order of magnitude as γ; therefore, we explore values of $\omega \in [0, 2]$. We observe that as ω increases, Ξ decreases and approaches zero; for $\omega \in [0.1, 1.6]$, the differences between the meshes are minimal, and the dependence on ω is weak ($\zeta = 0.25 \pm .01$ throughout).

Using these plots, we consider a few things when picking ω. Low flexibility and/or high stationarity would result in communities that are insensitive to real temporal fluctuations which inform the forecasting; the temporal edge weights dominate for $\omega > 1$. On the other end of the spectrum, high flexibility and/or low stationarity would mean that the strain loading is not being taken into account; the spatial edge weights dominate the communities for $\omega < 1$. Therefore, we select $\omega = 1$ because it balances the strain-loading history (inter-) with the rheological properties (intra-layer).

IV. RESULTS

We aim to determine whether we have identified the (known) geometry and location of Mud Creek with the results of the community detection algorithm (explored below). Once this is achieved, we examine the details of the degree of community persistence over time, and use our observations to construct a measure which is able to forecast increasing risk of failure.
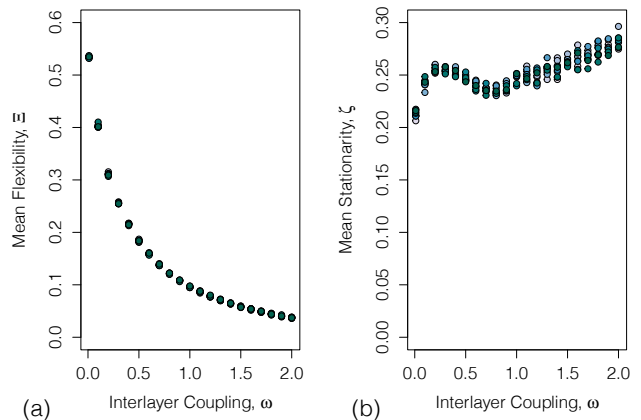

FIG. 4. (a) Network flexibility Ξ (Eq. 4) and (b) network stationarity ζ (Eq. 5) as a function of the choice of a particular temporal coupling parameter ω, for fixed spatial coupling $\gamma = 1$. Each data point is for one of the 10 meshes, distinguished by color.

A. Quantifying community structures

In order to identify areas of concern, we run community detection using the modularity function defined in Eq. 2 with parameters $\gamma = 1$ and $\omega = 1$. For each InSAR snapshot, there is a resulting community structure: 62 in total, 6 of which are shown in Fig. 5. We only present the results of one mesh in Fig. 5, but this is representative of the results for all 10 meshes, indicating that the algorithm is not sensitive to the particular mesh used. Mud Creek is identified as a community early on, clearly showing up in Fig. 5(c) in red and remains coherent for the next ~ 400 days prior to its ultimate failure.

We first consider a simple metric: how the number of communities found within each layer changes over time. In Fig. 5, the number of singleton communities (black nodes) increases as Mud Creek approaches failure, while the number of nonsingleton communities remains stable. Singleton communities most commonly appear at the border of communities; sometimes they join a community, and sometimes they arise at new boundaries when a community splits apart. As we will see later, these dynamics are connected to changes in persistence of community structures as Mud Creek approaches failure at $T = 0$.

To identify the underlying pattern within these identified communities, we visualize the evolution of the community structures in Fig. 6(a), where each horizontal line represents a single community, whose ID is assigned in order of its identification by the algorithm, and the color of the line represents the approximate location of that community (see inset).

There are two characteristics of a community that we considered: size and strength. The *strength* of a community is the mean of the weights of the edges, and the *size* of a community is the number of nodes. We find that a
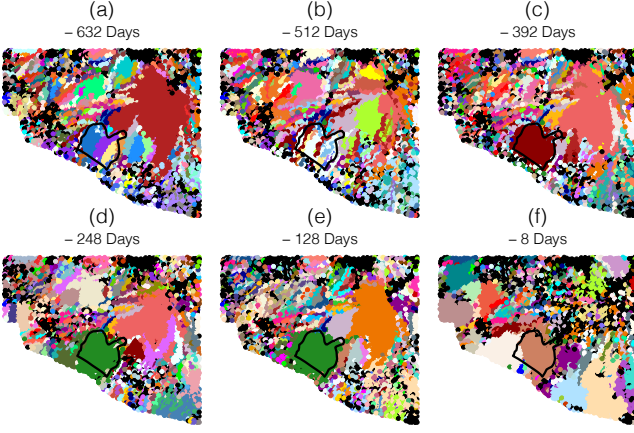
FIG. 5. Results of community detection using $\gamma = 1$, $\omega = 1$, shown for six representative layers prior to failure at $T = 0$ days. Nodes that are members of the same community have the same (arbitrary) color, corresponding to a single community ID number. Black nodes represent singleton communities. The black outline is Mud Creek.

joint measure of community size and strength, normalized by global values, is useful for highlighting the significance $\Psi_{c,\ell}$ of particular communities detected, calculated as

$$\Psi_{c,\ell} = \frac{n_{c,\ell}}{N} \frac{S_{c,\ell}}{S_\ell}, \tag{6}$$

where $n_{c,\ell}$ is the number of nodes in community c at layer ℓ, N is the total number of nodes, and $S_{c,\ell}$ and S_ℓ is the strength of community c and layer ℓ, respectively. Large values of $\Psi_{c,\ell}$ correspond to a community c being large, fast-moving, and/or steeply sloped.

Fig. 6 shows a graphical summary of the changes taking place in the overall community structure. Each horizontal bar represents a multilayer community during the range of dates (layers) it was detected, with the width proportional to $\Psi_{c,\ell}$. We observe 3 persistently strong communities, one each within Mud Creek, Region 1, and Region 2. Importantly, Mud Creek (outlined in the white polygon) is consistently detected as a strong community over a year before its failure, up through the final data collection. Within Region 2, the community we detect (outlined in the black polygon) corresponds to a creeping landslide that was not detected using InSAR alone; its detection here is an example of how community detection techniques can supplement more traditional ones. Finally, there is a community detected within Region 1 which intermittently appears as an area of concern, but has not been identified as a creeping landslide. Interestingly, no communities were identified within Region 3 because: while it shares similar topographic characteristics to Mud Creek, it is not creeping. Finally, we find that while this type of plot identifies areas of concern, it does not yet provide a framework for forecasting time of failure. In the next section, we develop such a metric.
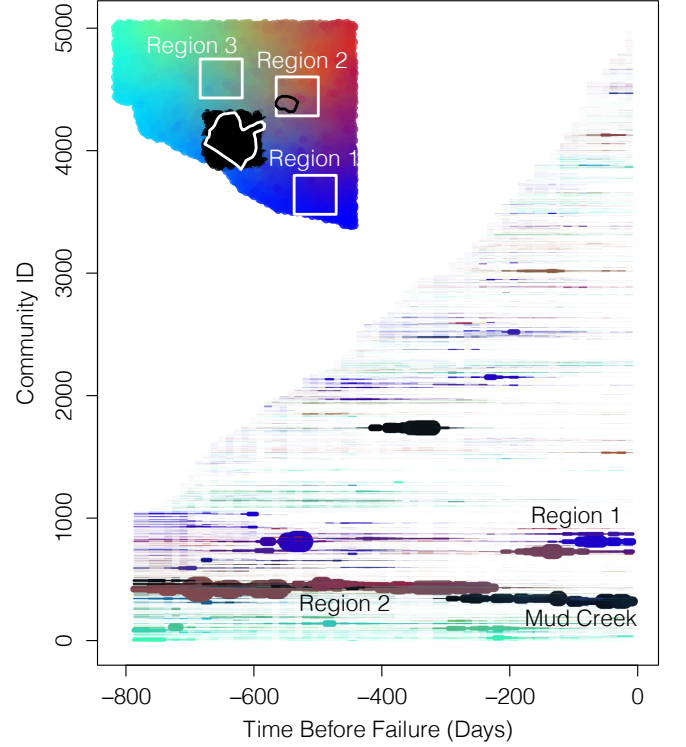

FIG. 6. Strip-chart showing the evolution of the community structures detected for parameters $\gamma = 1$ and $\omega = 1$, separately highlighting location, duration, and significance Ψ_c. Each detected community found has an ID number that is arbitrary (but corresponds to the order of detection), and is plotted as a line with three properties: color specifies its location (see inset map), the extent along the horizontal axis specifies its duration, and the thickness of of the line measures the size and strength of that community at layer ℓ, according to Eq. 6. The outlines of the creeping landslides found within the area are shown in the inset map.

B. Forecasting failure via community persistence

We observed that Mud Creek is identified as the same community for the last ~ 400 days in Fig. 6. Mud Creek is one of the few communities that clearly persists through time, while other communities split into smaller or singleton communities. Using these observations, we propose a new measure: *community persistence* Π_ℓ inspired by Rieck *et al.* [56], González-Avella *et al.* [57]. This measures the stability of the nodal composition for each community in relation to the community's size and is defined as

$$\Pi_\ell = \frac{1}{N} \sum_c \frac{c_{\ell-1} \cap c_\ell}{n_{c,\ell}}, \tag{7}$$

where N is the total number of nodes in the network, $n_{c,\ell}$ is the number of nodes in community c at layer ℓ, and $c_{\ell-1} \cap c_\ell$ is the number of nodes present in community c in both layers ℓ and $\ell-1$. An increase in community persistence indicates that a patch of nodes are consistently

being identified as a community c. This occurs when c is stronger-than-average for consecutive layers; if c were to grow or shrink, then its contribution to Π_ℓ is smaller than one. The more the communities change membership, the closer Π_ℓ is to zero.

In Fig. 7(a-b), we show Π_ℓ for varying values of ω from 0 to 2 at $\gamma = 1$ and varying values of γ from 0 to 2 at $\omega = 1$. Only near $\gamma = 1$ do we observe a signal using Π_ℓ, which confirms the common usage of this value [44, 54], and corresponds to communities that are stronger-than-average (*i.e.* indicative of a creeping hillslope). This happens because only a few communities are being picked up consistently with a similar subset of nodes from beginning to end. In Fig. 7(b), we obtain a similar result for values of $\omega \in [0, 2]$, but with more noise at low values of ω. Our choice of $\omega = 1$ captures the rising persistence without being hampered by temporal noise (discussed in §III B).To ensure that our analysis is not sensitive to the ensemble of 10 meshes that we use, we run community detection on 10 randomized meshes, and the mean for those 10 meshes is shown in Fig. 7.

We see a sharp increase in community persistence (shown in Fig. 7(c)) about 56 days before failure for all 10 meshes. This is indicative of a group of nodes, each one linked to spatial coordinates, that is continually being established as a community between each layer, in this case the nodes contained within the Mud Creek geometry. As this community becomes stronger (correlated with increasing velocity) and more persistent, we observe that the other communities start to break down into smaller communities, including the community representing the creeping landslide found within Region 2. This happens because the mean edge weight to partition nodes into communities (see Eq. 2) is increasing as we get closer to failure (reference Fig. 2).

V. DISCUSSION/CONCLUSION

We have developed an algorithm that takes a geospatial surface and represents it as a mesh of nodes. A few things we considered when creating the mesh are sensitivity to the mesh we use, artificial results, and good coverage of the area, all of which are addressed by using Poisson disc sampling. To test the spatial resolution of the DEM that is used, we compare a DEM with resolution of 12 m and the DEM we used for network analysis, as described in §II A, of a higher resolution of 1 m, and note that there are no significant differences between the two, as we downsample the 1 m DEM to 20 m.

We are able to model the dynamics and susceptibility of the system to identify Mud Creek both spatially (through community detection) and temporally (through the persistence of those communities). We do this by incorporating information about both the current state of rheology using velocity data derived from InSAR and the geometry of the area using slope. We found that using only velocity or slope was insufficient to
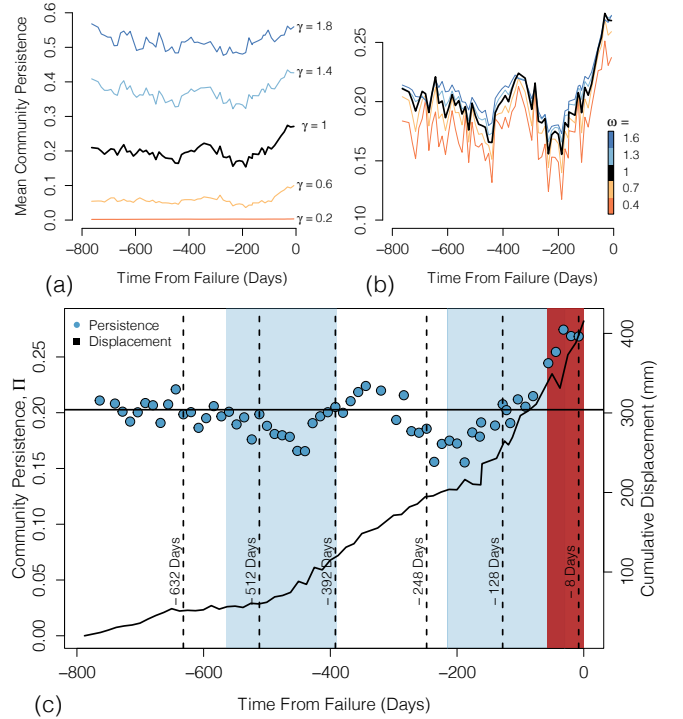

FIG. 7. Community persistence Π_ℓ (Eq. 7) as a function of time before failure, averaged over 10 meshes, (a) for different choices of spatial coupling parameter γ at fixed $\omega = 1$ and (b) for different choices of temporal coupling parameter ω at fixed $\gamma = 1.0$) (c) The mean community persistence Π_ℓ over 10 meshes using $\gamma = 1, \omega = 1$. Standard error, measured across the ensemble of 10 meshes, is smaller than the symbol size; the horizontal black line shows the temporal mean. The dashed lines represent the 6 snapshots of the community structures that are shown in Fig. 5. The blue shaded area represents the wet seasons shown in Fig. 1. The red shaded area highlights where Π_ℓ starts to significantly rise above the mean. The black line shows the maximum cumulative displacement within the Mud Creek polygon.

distinguish creeping landslides from the rest of the area: slope-weighted networks identified watershed basins and velocity-weighted networks identified large communities without separating them into geophysically-relevant areas. Through the inclusion of velocity and slope, we are able to pick up clusters of areas that are strongly-connected to each other. As shown in Fig. 6, we observe three areas of concern: one is Mud Creek, one is a creeping landslide that is already known, and one is a new area of interest for future monitoring. Combined with high-quality InSAR data, network science techniques has allowed us to forecast the location and geometry of the impending failure from 400 days out, without having to visually inspect a map (see Handwerger *et al.* [27]).

We quantify these results in Fig. 7; as we approach time of failure, there is a sharp increase in community persistence Π_ℓ. When compared to the maximum cumulative displacement of the landslide (shown in Fig. 7),

community persistence has a much clearer signal that time of failure is close, although it is not predictive enough to pinpoint a range of days. To test the success of this method using only the first wet season (during which Mud Creek did not fail), we repeated the full analysis using only the first 30 layers (up to -400 days). We correctly detected Mud Creek as an area of concern and correctly observed no sharp increase in Π_ℓ, which is consistent with the reality that failure was not imminent. We also note in Fig. 7 that the oscillations correspond to the wet/dry seasons, and that the method is not sensitive to choice of a specific mesh from the ensemble.

This raises several promising avenues for improvement in utility. From the standpoint of continually monitoring, it would be beneficial to instead use pixel tracking displacement or automatically processed InSAR data, instead of InSAR data that has been manually corrected, such as the work of Handwerger et al. [27]. Previous examination of the InSAR time series by Handwerger et al. [27] reveal that Mud Creek exhibited strong seasonal displacement driven by precipitation. The displacement increased over the study period due to increasing rainfall during the study period. Although the landslide behavior changed from year to year, and the landslide was moving notably faster in 2017 than previous years, it was concluded by Handwerger et al. [27] that there was no displacement pattern that is indicative of a slope approaching runaway instability. However, we note here that this is likely a result of bias in the InSAR data itself. InSAR cannot detect large displacements ($>$ half the radar wavelength), which means that if the slope is moving too fast between acquisitions there can be phase bias and an underestimate of true displacement. Warrick et al. [31] showed that the landslide moved over 10 m between March and May 2017, which is too large of displacement for InSAR to measure. From this, we know that the InSAR-measured displacement over the final months prior to collapse is much too slow. Therefore, the quantities used are a lower bound, and improved data would provide for an even stronger signal than is reported here. Significant outlay of human resources would be required to avoid underestimating ground displacements and being dominated by noise, and it remains to be understood what level of spatial, temporal, and velocity resolution is required to achieve the type of results we observed here.

The inclusion of time-delayed rainfall layers, or even approximations to soil moisture (proxy for pore-water pressure) would likely improve the fidelity of the multilayer network in describing the evolving state of the study site. Furthermore, the topography of Earth's surface is constantly changing due to ground deformation, so using a static DEM may introduce challenges and loss of crucial information over a long time period.

While community detection improves the signals of forecasting failure in the Mud Creek area and aids in the identification of creeping landslides, an important open question is how does this method fare on other areas, as well as how to improve the temporal signal of failure.

VI.   ACKNOWLEDGEMENT

[1] P. Lacroix, A. L. Handwerger, and G. Bièvre, Nature Reviews Earth & Environment **1**, 404 (2020).

[2] X. Hu, R. Bürgmann, W. H. Schulz, and E. J. Fielding, Nature Communications **11**, 2792 (2020).

[3] E. Rutter and S. Green, Journal of the Geological Society **168**, 359 (2011).

[4] A. Aryal, B. A. Brooks, M. E. Reid, G. W. Bawden, and G. R. Pawlak, Journal of Geophysical Research: Earth Surface **117**, 10.1029/2011JF002161 (2012).

[5] A. Tordesillas, S. Kahagalage, L. Campbell, P. Bellett, E. Intrieri, and R. Batterham, Scientific Reports **11**, 9729 (2021).

[6] T. Carlà, E. Intrieri, F. Raspini, F. Bardi, P. Farina, A. Ferretti, D. Colombo, F. Novali, and N. Casagli, Scientific Reports **9**, 1 (2019).

[7] E. Intrieri, T. Carlà, and G. Gigli, Earth-Science Reviews **193**, 333 (2019).

[8] K. Singh and A. Tordesillas, Entropy **22**, 67 (2020).

[9] A. Tordesillas, Z. Zhou, and R. Batterham, Mechanics Research Communications **92**, 137 (2018).

[10] L. Cascini, M. R. Scoppettuolo, and E. Babilio, Landslides 10.1007/s10346-022-01934-3 (2022).

[11] A. K. Gabriel, R. M. Goldstein, and H. A. Zebker, Journal of Geophysical Research **94**, 9183 (1989).

[12] G. Franceschetti and R. Lanari, *Synthetic Aperture Radar Processing*, 1st ed. (CRC Press, 2018).

[13] F. Di Traglia, C. De Luca, M. Manzo, T. Nolesini, N. Casagli, R. Lanari, and F. Casu, Remote Sensing of Environment **260**, 112441 (2021).

[14] K. Teshebaeva, H. P. Echtler, B. Bookhagen, and M. R. Strecker, Earth Surface Processes and Landforms , esp.4648 (2019).

[15] K. Teshebaeva, S. Roessner, H. Echtler, M. Motagh, H.-U. Wetzel, and B. Molodbekov, Remote Sensing **7**, 8973 (2015).

[16] T. Carlà, P. Farina, E. Intrieri, H. Ketizmen, and N. Casagli, Engineering Geology **235**, 39 (2018).

[17] M. S. Bernardi, P. C. Africa, C. de Falco, L. Formaggia, A. Menafoglio, and S. Vantini, Mathematical Geosciences **53**, 1781 (2021).

[18] G. J. Dick, E. Eberhardt, A. G. Cabrejo-Liévano, D. Stead, and N. D. Rose, Canadian Geotechnical Journal **52**, 515 (2015).

[19] S. Zhou, A. Tordesillas, E. Intrieri, F. Di Traglia,

G. Qian, and F. Catani, Journal of Geophysical Research: Solid Earth **127**, 10.1029/2021JB022957 (2022).

[20] S. Das and A. Tordesillas, Remote Sensing **11**, 2777 (2019).

[21] F. Agliardi, M. M. Scuderi, N. Fusi, and C. Collettini, Nature Communications **11**, 1352 (2020).

[22] L. Torres, A. S. Blevins, D. Bassett, and T. Eliassi-Rad, SIAM Review **63**, 435 (2021).

[23] A. Le Bouil, A. Amon, S. McNamara, and J. Crassous, Physical Review Letters **112**, 246001 (2014).

[24] F. Fazelpour, *Failure and Flow in Sheared Granular Materials*, PhD Thesis, North Carolina State University (2022).

[25] J. Mei, G. Ma, Q. Wang, T. Wu, and W. Zhou, International Journal of Solids and Structures **251**, 111763 (2022).

[26] A. Tordesillas, Philosophical Magazine **87**, 4987 (2007).

[27] A. L. Handwerger, M.-H. Huang, E. J. Fielding, A. M. Booth, and R. Bürgmann, Scientific Reports **9**, 1569 (2019).

[28] J. J. Roering, B. H. Mackey, A. L. Handwerger, A. M. Booth, D. A. Schmidt, G. L. Bennett, and C. Cerovski-Darriau, Geomorphology **236**, 109 (2015).

[29] A. L. Handwerger, J. J. Roering, D. A. Schmidt, and A. W. Rempel, Geomorphology **246**, 321 (2015).

[30] J. S. Scheingross, B. M. Minchew, B. H. Mackey, M. Simons, M. P. Lamb, and S. Hensley, Geological Society of America Bulletin **125**, 473 (2013).

[31] J. A. Warrick, A. C. Ritchie, K. M. Schmidt, M. E. Reid, and J. Logan, Landslides **16**, 1201 (2019).

[32] PRISM Climate Group, Oregon State U (2015), accessed on 2022-01-13 `https://prism.oregonstate.edu/`.

[33] D. L. Swain, B. Langenbrunner, J. D. Neelin, and A. Hall, Nature Climate Change **8**, 427 (2018).

[34] R. J. Allen and R. Luptowitz, Nature Communications **8**, 16055 (2017).

[35] R. M. Iverson, M. E. Reid, N. R. Iverson, R. G. LaHusen, M. Logan, J. E. Mann, and D. L. Brien, Science **290**, 513 (2000).

[36] N. Mitarai and F. Nori, Advances in Physics **55**, 1 (2006).

[37] W. H. Schulz, J. P. McKenna, J. D. Kibler, and G. Biavati, Landslides **6**, 181 (2009).

[38] N. Lu and J. Godt, *Hillslope Hydrology and Stability* (Cambridge University Press, Cambridge, 2013).

[39] L. Girard, D. Amitrano, and J. Weiss, Journal of Statistical Mechanics: Theory and Experiment **2010**, P01013 (2010).

[40] J. C. Jaeger, N. G. W. Cook, and R. W. Zimmerman, *Fundamentals of rock mechanics*, 4th ed. (Blackwell Pub, Malden, MA, 2007).

[41] G. Palla, A.-L. Barabási, and T. Vicsek, Nature **446**, 664 (2007).

[42] K. A. Whitaker, Z. Varga, L. C. Hsiao, M. J. Solomon, J. W. Swan, and E. M. Furst, Nature Communications **10**, 2237 (2019).

[43] S. Fortunato, Physics Reports **486**, 75 (2010), arXiv: 0906.0612.

[44] L. Papadopoulos, J. G. Puckett, K. E. Daniels, and D. S. Bassett, Physical Review E **94**, 032908 (2016).

[45] U. S. Geological Survey (2014), ARRA-CA CentralCoast-Z4 2010, `https://portal.opentopography.org/usgsDataset?dsid=ARRA-CA_CentralCoast-Z4_2010`, last accessed on 2021-05-17.

[46] R. L. Cook, ACM Transactions on Graphics **5**, 51 (1986).

[47] M. A. Porter, J.-P. Onnela, and P. J. Mucha, Notices of the American Mathematical Society **56**, 1082 (2009).

[48] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, Science **328**, 876 (2010).

[49] M. E. J. Newman and M. Girvan, Physical Review E **69**, 026113 (2004).

[50] M. E. J. Newman, Proceedings of the National Academy of Sciences **103**, 8577 (2006).

[51] L. G. S. Jeub, M. Bazzi, I. S. Jutla, and P. J. Mucha, A generalized louvain method for community detection implemented in matlab (2011-19), downloaded on 2020-05-15 `http://netwiki.amath.unc.edu/GenLouvain`, `https://github.com/GenLouvain/GenLouvain`.

[52] M. Sarzynska, E. A. Leicht, G. Chowell, and M. A. Porter, Journal of Complex Networks **4**, 363 (2016).

[53] D. S. Bassett, E. T. Owens, M. A. Porter, M. L. Manning, and K. E. Daniels, Soft Matter **11**, 2731 (2015).

[54] D. S. Bassett, N. F. Wymbs, M. A. Porter, P. J. Mucha, J. M. Carlson, and S. T. Grafton, Proceedings of the National Academy of Sciences **108**, 7641 (2011).

[55] D. S. Bassett, M. A. Porter, N. F. Wymbs, S. T. Grafton, J. M. Carlson, and P. J. Mucha, Chaos: An Interdisciplinary Journal of Nonlinear Science **23**, 013142 (2013).

[56] B. Rieck, U. Fugacci, J. Lukasczyk, and H. Leitte, IEEE Transactions on Visualization and Computer Graphics **24**, 822 (2018).

[57] J. C. González-Avella, M. G. Cosenza, J. L. Herrera, and K. Tucci, EPL (Europhysics Letters) **107**, 28002 (2014).

[58] Numerical methods are published at V. D. Desai, networkLandslide, `https://github.com/vddesai-97/networkLandslide`
.