DeFiNe: an optimisation-based method for robust disentangling of filamentous networks

David Breuer[1,*] and Zoran Nikoloski[1]

[1]*Systems Biology and Mathematical Modeling, Max Planck Institute of Molecular Plant Physiology,*
Am Muehlenberg 1, 14476 Potsdam, Germany and
[*]*breuer@mpimp-golm.mpg.de*

Thread-like structures are pervasive across scales, from polymeric proteins to root systems to galaxy filaments, and their characteristics can be readily investigated in the network formalism. Yet, network links usually represent only parts of filaments, which, when neglected, may lead to erroneous conclusions from network-based analyses. The existing alternatives to detect filaments in network representations require tuning of parameters over a large range of values and treat all filaments equally, thus, precluding automated analysis of diverse filamentous systems. Here, we propose a fully automated and robust optimisation-based approach to detect filaments of consistent intensities and angles in a given network. We test and demonstrate the accuracy of our solution with contrived, biological, and cosmic filamentous structures. In particular, we show that the proposed approach provides powerful automated means to study properties of individual actin filaments in their network context. Our solution is made publicly available as an open-source tool, "DeFiNe", facilitating decomposition of any given network into individual filaments.

Keywords: polymers, cytoskeleton, networks, path cover, computational complexity

INTRODUCTION

Many network-like structures in nature are composed of filaments forming intricate interconnected arrays across different scales of organisation. For instance, filamentous structures can be observed in networks of cellulose polymers in the primary cell wall of plants and algae [59, 103], cytoskeletal networks of actin filaments or microtubules in cells across all domains of life [71, 96, 116], networks of neurons [19, 66], root systems [38, 72, 124], as well as solar prominences [42, 74] and galaxy clusters [16, 17, 104, 110]. Network-based studies of these structures have already elucidated important aspects such as the mechanics of cellulose networks [81, 103], transport on cytoskeletal actin networks [2, 9], and connectivity patterns in the brain [55, 66, 101]. However, the network links usually correspond to segments of the filaments; therefore, the classical network-based analysis neglects the identities of individual filaments. A few powerful exceptions have recently started to emerge [120, 121] which may identify multiple segments that belong to the same filament; yet, since these studies do not capture filament overlaps, filaments are still broken into potentially multiple fragments. Characterisation of the mechanical- [12, 61, 73], transport- [9, 85], and information-transmission related properties [13, 32] in such network representations may hence lead to erroneous conclusions due to their differences within and between filaments. Thus, analysis of filamentous structures rests upon accurate identification of individual filaments.

Since most of the filamentous structures in natural and man-made systems are studied by using imaging technologies, filaments are identified either directly from the imaging data or from networks extracted from these data (see Tab. 1 for succinct review). In the first class of approaches, a texture-based method is employed to in-

fer the overall orientation of objects in an image section [18]. However, this method cannot be employed to pinpoint individual filaments. Another method decomposes entire images of filamentous structures into linear segments based on a linear programming formulation [118]. While this method utilises few parameters (e.g., number of filaments), it only models and extracts a representative set of linear filaments. Moreover, filaments have been modelled as linear segments, detected by co-localisation with a parallel grid at different orientations and by using manually chosen intensity thresholds along a filament [54]. While this method is fast and useful for extracting linear filaments (e.g., microtubules), it does not capture bent or tangled filaments and necessitates manual parameter selection. Alternatively, tracing- and tracking-based methods which start from one or multiple image points and predict neighbouring points on a putative filament through optimisation of an energy function are powerful methods for filament identification. Although these algorithms have led to great insights, especially into the connectome, they typically require user input and do not capture overlapping filaments [27, 76, 77, 88]. Using a similar approach, open contour-based methods employ deformable curve models that elongate and align according to an energy functional to match the target filaments. Recent advances in open contour-based approaches enable fully automated filament detection [120, 121], but can account for the overlap of only few filaments at the expense of parameter tuning [97].

The second class of approaches for disentangling filamentous structures employs a two-step procedure: First, weighted networks are extracted from image data from different systems and imaging sources. There is a large variety of algorithms for this task [11, 27, 83, 89] which vary, in particular, in the number of parameters. Some of the methods from the first class, presented above, may also be used to obtain such network representations

Input	Method	Features					References
		curved filaments	filament -specific	intensity -based	automated	parsi- monious	
image	texture filter	−	−	+	+	+	[18]
	linear programming	−	−	+	+	+	[118]
	rotating grid	−	+	+	+	+	[54]
	filament tracing	+	+	+	○	○	[27, 77, 88]
	filament tracking	+	+	+	+	○	[76]
	open contours	+	+	+	+	○	[97, 121]
network	rule-based decomp.	+	+	−	○	○/−	[65, 89]
	filament cover	+	+	+	+	○/+	current work

Table 1. **Overview of different approaches for disentangling filamentous networks.** Two main classes of approaches to analyse the filamentous structure of networks can be distinguished, based on whether they operate on image data or on extracted networks. Irrespective of the class, the existing approaches vary in their capacity (+) or inability (−) to detect curved filaments, identify individual filaments, and to include information about the intensity/thickness of filaments. Further, the amount of manual user input as well as the number of parameters required by the algorithms can be feasible (+), laborious (−), or depends on the specific variant of the algorithm (○). For the network-based approaches, the number of required parameters may be different for the extraction of the network from image data and the consequent decomposition of the network into filaments (separated here by /).

(e.g. [76, 121]). Second, the given, weighted networks are decomposed into filaments. The two existing methods for this task [65, 89] define specific junctions for bifurcations and crossings of filaments, depending on the distances between nodes, and assign filament identities according to manually chosen angle thresholds between incoming and outgoing edges. In particular, they strongly restrict the potential overlap of filaments and, due to the angle constraints, allow only crossing but no touching filaments. Most importantly, these methods require manual parameter selection and do not take into account filament intensity/thickness. We note that the step of decomposing a given network may also be beneficially applied to networks obtained, e.g., by open contour-based approaches in which filaments have been fragmented due to omission of filament overlaps [120, 121].

Here, we propose a robust approach to decompose a weighted network into an optimal set of individual filaments. Therefore, our approach addresses the second step in the second class of approaches, presented above. The decomposition is based on a computationally difficult problem, referred to as filament cover problem (FCP), for which we propose suitable approximation algorithms. We test and demonstrate the accuracy of the findings from the approximation algorithms on artificial as well as biological and cosmic filamentous networks by comparison to manually obtained filament covers. In addition, we demonstrate that the proposed, fully automated solution allows facile characterisation of well-studied properties of individual filaments, for which alternative approaches require parameter tuning or time-consuming manual tracing. The proposed approach is implemented in a publicly available open-source tool, "DeFiNe" (**De**composing **Fi**lamentous **Ne**tworks), which can be used to decompose any given weighted network into a set of individual filaments for further analyses (http://mathbiol.mpimp-golm.mpg.de/DeFiNe/).

METHODS

In this section we introduce the mathematical formulation of our optimisation-based approach to decompose filamentous networks, demonstrate its computational intractability, and formulate a suitable approximation scheme. Moreover, we introduce new quality measures which take into account the underlying network structures for the comparison of the obtained filament decompositions with manual assignments used as a gold standard. Finally, we provide a brief overview of the studied data from different biological and physical systems. While we believe that these more technical explanations may promote a deeper understanding of our and related approaches, we encourage readers familiar with the aforementioned topics to proceed directly to the Results.

Mathematical formulation of the filament cover problem

Any filamentous structure may be represented as weighted geometric graph $G = (\mathcal{N}, \mathcal{E})$ with $N = |\mathcal{N}|$ nodes and $E = |\mathcal{E}|$ undirected, weighted edges. Edges represent filament segments and their intensities or thicknesses are reflected by their weights w_e, $e := (n_0, n_1) \in \mathcal{E}$ and $n_0, n_1 \in \mathcal{N}$. Nodes represent endpoints of filament segments and their positions are denoted by v_n, $n \in \mathcal{N}$, whereby, typically, $v_n \in \mathbb{R}^2$ or $v_n \in \mathbb{R}^3$ for networks extracted from image data.

We naturally represent a filament by an edge-path, $p = (e_{p,1}, \ldots, e_{p,P})$, $e \in \mathcal{E}$, i.e., by an ordered sequence of $P = |p|$ adjacent edges, where $e_{p,i}$ denotes the i-th edge of filament p. The quality of a given filament p is assessed

by the pairwise filament roughness

$$r_{p,\text{pair}} = \begin{cases} (P-1)^{-1} \sum_{i=1}^{P-1} \left| w_{e_{p,i+1}} - w_{e_{p,i}} \right| & , P > 1 \\ w_{e_{p,1}} & , P = 1 \end{cases}, (1)$$

where $w_{e_{p,i}}$ denotes the weight of the i-th edge in filament p. The pairwise filament roughness is small if the edge weights along a filament vary smoothly, as expected for natural filaments (but cf. Discussion). For filaments that consist of one edge only, their roughness is given by their edge weight to increase the flexibility of our approach (cf. Supplemental Material S1). Other roughness measures may be readily introduced that take into account filament thicknesses or alignments. As an additional example, we study the all-to-all filament roughness

$$r_{p,\text{all}} = \begin{cases} (P-1)^{-1} \max_{i,j \in \{1,...,P\}} \left| w_{e_{p,i}} - w_{e_{p,j}} \right| & , P > 1 \\ w_{e_{p,1}} & , P = 1 \end{cases}, (2)$$

which is the average maximal difference between any two edge weights in a filament p, and again the original weight of the edge is used for a filament of length one. Taking into account that most filaments are only moderately bent, we further consider the maximal filament deflection angle between adjacent edges of a path p,

$$r_{p,\text{angle}} = \max_{i \in \{1,...,P-1\}} \quad (3)$$
$$\text{angle}\left(v_{e_{p,i+1,1}} - v_{e_{p,i+1,0}}, v_{e_{p,i,1}} - v_{e_{p,i,0}} \right)$$

where $v_{e_{p,i,0}}$ and $v_{e_{p,i,1}}$ denote the positions of the start and end nodes of the i-th edge of filament p, respectively. Moreover, $\text{angle}\left(v, v' \right) := \arccos\left(\frac{v \cdot v'}{\sqrt{v \cdot v}\sqrt{v' \cdot v'}} \right)$ is the Euclidean angle of two vectors v and v' and $r_{p,\text{angle}} = 0°$ corresponds to perfectly straight alignment.

The optimal decomposition of a network into individual, smooth filaments then corresponds to solving our filament cover problem (FCP; cf. Supplemental Material S1 for an overview of related cover problems):

Given a graph $G = (\mathcal{N}, \mathcal{E})$ and the set $\mathcal{P}$ of all edge-paths in G with roughnesses r_p, $p \in \mathcal{P}$:

Find a subset $\mathcal{P}_{\text{fil}} \subseteq \mathcal{P}$ with minimal total (or average) roughness R such that each element in $\mathcal{E}$ is covered (at least) once.

Here, edges that are covered by more than one path naturally correspond to filament overlaps. Minimising the average instead of the total roughness yields shorter filaments, as appropriate for some networks (cf. Supplemental Material S1).

Computational intractability of the filament cover problem and approximation algorithm

The FCP is computationally intractable on general and even planar graphs (cf. Supplemental Material S2 for motivation and proof). Graphs generated from two-dimensional image data are planar by construction [11, 83]. The proof is by reduction from the well-studied Hamiltonian path problem which asks, for a given network, whether there is a sequence of adjacent nodes that includes each node exactly once, and which is known to be intractable on planar graphs [41]. Moreover, we outline an algorithm for solving the FCP in polynomial time on trees (cf. Supplemental Material S3).

Since the FCP is computationally intractable on general and even planar graphs, we devise an approximation scheme by formulating the FCP as a fractional integer linear program (cf. Supplemental Material S4 for motivation and details). For a given set $\mathcal{P}' \subseteq \mathcal{P}$ of input paths with pairwise filament roughnesses r_p, $p \in \mathcal{P}'$, we solve:

$$\text{minimize} \frac{\sum_{p \in \mathcal{P}'} r_p x_p}{\left(\sum_{p \in \mathcal{P}'} x_p \right)^A} \quad (4)$$
$$\text{subject to} \sum_{p:e \in p} x_p \geq 1 \text{ for all } e \in \mathcal{E}$$
$$x_p \in \{0,1\} \text{ for all } p \in \mathcal{P}',$$

where we use $r_p \in \{r_{p,\text{pair}}, r_{p,\text{all}}\}$ (Eqs. 1 or 2; referred to as *pair* and *all*). In the first line, $A \in \{0,1\}$ determines whether the total or the average roughness is minimised (*total/avg*). The inequality in the second line allows overlapping filaments and equality holds for an exact cover (*over/exact*). For $A = 0$, Eq. 4 is a binary linear program and for $A = 1$, the fractional problem Eq. 4 may be rewritten as a binary linear program (cf. Supplemental Material S4). Binary linear programs may be solved using well-established and efficient algorithms [70, 94].

To solve the FCP for a given network, we further need to collect a set of input paths $\mathcal{P}' \subseteq \mathcal{P}$. Since for a general graph it is not feasible to collect all paths $\mathcal{P}$ (cf. Supplemental Material S2), we propose two approaches (referred to as *RMST* and *BFS*): (1) We create $T = 100$ random minimal spanning trees (RMST) of G whose $N(N-1)/2$ non-trivial, undirected paths are added to our set $\mathcal{P}'$. To obtain a RMST, each edge is assigned a randomly and uniformly distributed weight and the minimum spanning tree with respect to these weights is computed. (2) We perform a modified breadth-first search (BFS) on the nodes, stop the search for a path p when it violates the straightness criterion $r_{p,\text{angle}} < 60°$ (cf. Eq. 3), and add all permitted paths to $\mathcal{P}'$. We note that for real-world filamentous graphs, the number of nodes and their degrees are constrained by the filament thickness, while the number of considered loops is further restricted by the straightness criterion, so that our heuristically modified BFS yields a representative set $\mathcal{P}'$ of paths in reasonable time. Moreover, we note that the 60°-criterion is introduced for computational reasons and provides a tolerant estimate for maximal bending of the studied real-world filaments which are typically less bent.

Quality assessment of filament covers via structure-aware partition similarity measures

The accuracy of the filaments covers obtained by solving the FCP is assessed by comparison to manual filament assignments (cf. Fig. 1b). We quantify the similarity of the two partitions of the set of edges into (potentially overlapping) filaments using the variation of information, VI, the Jaccard index, JI, and the Rand index, RI,

$$\mathrm{VI}\left(\mathcal{C},\mathcal{C}'\right) = 1 + (U \log U)^{-1} \cdot \qquad (5)$$

$$\cdot \sum_{i,j} g_{i,j} \left(\log\left(\frac{g_{i,j}}{g_{\cdot,j}} \right) + \log\left(\frac{g_{i,j}}{g_{i,\cdot}} \right) \right),$$

$$\mathrm{RI}\left(\mathcal{C},\mathcal{C}'\right) = \frac{h_{=,=} + h_{\neq,\neq}}{h_{=,=} + h_{=,\neq} + h_{\neq,=} + h_{\neq,\neq}}, \qquad (6)$$

$$\mathrm{JI}\left(\mathcal{C},\mathcal{C}'\right) = \frac{h_{=,=}}{h_{=,=} + h_{=,\neq} + h_{\neq,=}}, \qquad (7)$$

where $U = \sum_{i=1}^{C} |\mathcal{C}_i| = \sum_{j=1}^{C'} \left| \mathcal{C}'_j \right|$, $g_{i,j} = \left| \mathcal{C}_i \cap \mathcal{C}'_j \right|$, $g_{\cdot,j} = \sum_{i=1}^{C} g_{i,j}$, and $g_{i,\cdot} = \sum_{j=1}^{C'} g_{i,j}$ [31, 79, 93]. The contingency tables $h_{\times,\times'}$, $\times, \times' \in \{=, \neq\}$, provide the numbers of edge pairs which are in the same or different sets in the two partitions, respectively. While these classical measures are widely used [63, 79], they may generally yield opposing results and VI is not well-defined for overlapping partitions (cf. Supplemental Material S6). More severely, these measures do not take into account the structure of the graph underlying the partitions. To remedy this shortcoming, we introduce a suite of measures, the structure-aware Rand and Jaccard indices (cf. Eqs. 6 and 7),

$$\mathrm{RI}^d\left(\mathcal{C},\mathcal{C}'\right) = \frac{h^d_{=,=} + h^d_{\neq,\neq}}{h^d_{=,=} + h^d_{=,\neq} + h^d_{\neq,=} + h^d_{\neq,\neq}}, \qquad (8)$$

$$\mathrm{JI}^d\left(\mathcal{C},\mathcal{C}'\right) = \frac{h^d_{=,=}}{h^d_{=,=} + h^d_{=,\neq} + h^d_{\neq,=}}. \qquad (9)$$

Here $h^d_{\times,\times'}$, $\times, \times' \in \{=, \neq\}$, $d \in \mathbb{N}_{>0}$, count the number of edge pairs which are in the same or different sets in the two partitions and which are separated by at most d nodes in G (cf. Supplemental Material S6 for details). Thus, RI^1 and JI^1 yield structure-aware measures of partition similarity that consider only the partition memberships of adjacent edges (local perspective), while $\mathrm{RI}^\infty \equiv \mathrm{RI}$ and $\mathrm{JI}^\infty \equiv \mathrm{JI}$ do no not take into account the positions of edges in the graph and reproduce the original measures (global perspective; cf. Supplemental Material S6 for an extensive comparison of similarity measures and intermediates between local and global perspective).

Extraction of weighted networks from image data

We test our method to disentangle filamentous networks on various weighted, geometric networks extracted from image data. The network extraction procedure is similar to those proposed in [11, 83] (cf. Supplemental Material S5 for details). We analyse (1) two artificial networks extracted from drawn filamentous patterns, (2) two cytoskeletal networks from confocal microscope images of *Arabidopsis thaliana* hypocotyl actin cytoskeletons [22], (3) 100 additional cytoskeletal networks from a movie over 200 s from the same experimental setup, (4) two neural networks from a fluorescence microscopy image of a branching rat hippocampal neuron *in vitro* [20] and a schematic of a cat retinal ganglion cell [75], respectively, and (5) two cosmic networks obtained from images of simulated galaxy clusters [104] (see Tab. 2 for an overview).

RESULTS

Decomposing filamentous networks is a hard optimisation problem

A filamentous network is naturally represented as a weighted graph, whereby the links (i.e., edges) denote segments of filaments and the nodes represent the ends of the segments. The edge weights typically capture the intensity or thickness of the filament segments. In this network representation, a filament corresponds to a path given by an ordered sequence of adjacent edges. To identify individual filaments, we seek a decomposition of the set of edges into paths so that each edge is covered (i.e., belongs to at least one path). Edges belonging to more than one path naturally model filament overlaps. We will refer to such a decomposition as a filament cover. Since a filament cover is non-unique, we introduce a quality measure, called roughness, to assess the quality of each path and the cover itself. Here we mainly consider the pairwise filament roughness given by the average absolute value of weight differences between adjacent edges. This roughness measure quantifies how strongly the thickness varies along a filament and is typically small for biological filaments. Disentangling the filamentous network amounts to solving the filament cover problem (FCP): Find a set of paths of minimum sum of roughness values that covers the network (cf. Methods and Supplemental Material S1 for the mathematical formulation). The FCP formulation is quite versatile: For instance, instead of minimising the total roughness of the filament cover, we may minimise the average roughness. This optimisation objective favours shorter filaments and may be more appropriate for specific types of networks. Other roughness measures (e.g., considering the spatial alignment of edges to penalise filaments with strong curvature) are readily introduced and can be considered in a multi-objective optimisation approach

(cf. Methods and Supplemental Material S1 for different measures).

While providing a well-defined approach towards disentangling filamentous networks, solving the FCP is computationally prohibitive. Indeed, we show that the FCP is intractable even on planar graphs (cf. Methods and Supplemental Material S2) which are used to represent filamentous structures extracted from 2D image data [11, 83]. While the FCP is solvable in polynomial time on trees (cf. Supplemental Material S3), most biological filamentous structures are not tree-like as they contain loops [22, 58, 83]. Therefore, we propose suitable approximation schemes to the FCP for the considered networks (cf. Methods and Supplemental Material S4 for details and the mathematical formulation). The approximation schemes rely on collecting a large sample of paths in a given graph, followed by the computation of the roughness of each path. The paths are collected by performing a modified breadth-first search (BFS) or by sampling from random minimum spanning trees (RMST). Next, we write the FCP as classical set cover problem [56] which aims at covering the set of edges with a subset of the collected paths of minimum total or average roughness. The set cover approximation of FCP can be formulated and solved as a (fractional) binary linear program for which well-established algorithms exist [94]. The output of the program is a set of paths which correspond to the individual filaments of the studied network. Summarising, the FCP may be solved with different options: The initial set of paths is obtained from a modified BFS (denoted by *BFS*) or sampling of RMSTs (*RMST*), the filaments may overlap (*over*) or not (*exact*), a pairwise (*pair*) or all-to-all filament roughness measure (*all*) is used, and the total (*total*) or average (*avg*) roughness is minimised. Since all these options are categorical, all possible $2^4 = 16$ combinations may be readily checked and no data-specific and computationally demanding gauging of continuous parameters is necessary, as is the case for related approaches [65, 89]. We provide an open-source implementation of our approach, termed "DeFiNe" (**De**composing **Fi**lamentous **Ne**tworks), with a simple and user-friendly graphical user interface available at http://mathbiol.mpimp-golm.mpg.de/DeFiNe/. DeFiNe takes as input a weighted graph in the standard .gml file format [50] and outputs a .gml graph with filament identities stored as edge colours as well as a standard, human-readable .csv-table of various individual filament measures for custom analyses.

Disentangling artificial filamentous structures

To test the accuracy of our approach, we investigate an artificial network (Fig. 1a) of pre-specified filamentous structure (Fig. 1b; cf. Methods and Supplemental Material S4 for the extraction of the network; cf. Supplemental Material S9 for an overview of the different stages of our approach, from an images to a network to fil-

aments). The network contains crossing and overlapping filaments as well as a loop (Fig. 1b, $\otimes$, $\ominus$, and $\odot$, respectively). First, we automatically decompose the weighted filamentous network by solving the FCP for a set of input paths from a modified BFS, allowing for overlaps, using the pairwise roughness measure, and minimising the total roughness of the cover (Fig. 1c, cf. Eq. 4). The filament identities and colours are matched by solving an assignment problem (cf. [60, 117]) such that the total number of edges shared by two filaments, from the manual assignment and the automated cover, is maximised. The agreement between the automated cover and the manual assignment may be measured by classical partition similarity measures such as the Jaccard index JI which counts the fraction of edge pairs which are part of same filament [31, 79]. However, JI does not take into account the structure of the underlying network. Hence, we introduced a new similarity measure, JI^1, that considers only pairs of adjacent edges in each filament and thus incorporates the network structure (cf. Methods and and Supplemental Material S6 for details, a generalisation to JI^d that considers only pairs of edges which are separated by at most d nodes, and a comparison of various similarity measures). For our artificial network, solving the above FCP yields a decomposition which agrees excellently with the manual assignment ($JI = JI^1 = 1$) as all filaments are correctly detected. Second, we choose a different set of input paths obtained from sampling RMSTs for solving th FCP (Fig. 1d). While most filaments are correctly detected, the loop (cf. Fig. 1b) is over-segmented ($\oplus$) because it is not contained in the set of input paths in its entirety (due to looplessness of trees). Third, we solve the exact FCP which does not allow overlapping filaments (Fig. 1e). Expectedly, the agreement with the manual assignments is lower because filaments are over-segmented into disjoint segments and the supposedly overlapping parts are under-segmented ($\ominus$), i.e., the respective edges are assigned to a single filament instead of multiple filaments. Finally, we employ the all-to-all roughness measure to assess the quality of the filaments (Fig. 1f, cf. Eq. 2). Filament crossings, overlaps, and the loop are again correctly detected but parts of two filaments are interchanged (cf. $\oplus$). This is due to the intensity/thickness of the underlying filaments which is consistently higher for the new detected filaments which are therefore favoured by the all-to-all roughness measure. These test cases demonstrate the versatility and the accuracy of the proposed approach to decompose a given network into filaments.

In the analysis of many real-world filamentous structures, the knowledge of the underlying network structure is incomplete or the image data impede filament detection due to low signal-to-noise ratios. To investigate the effect of these obstacles on robust filament detection, we study two scenarios (Supplemental Material S7): In the first scenario, we remove a single edge from the network, recompute the optimal filament cover, and calculate its agreement with the manual filament assignment as meas-
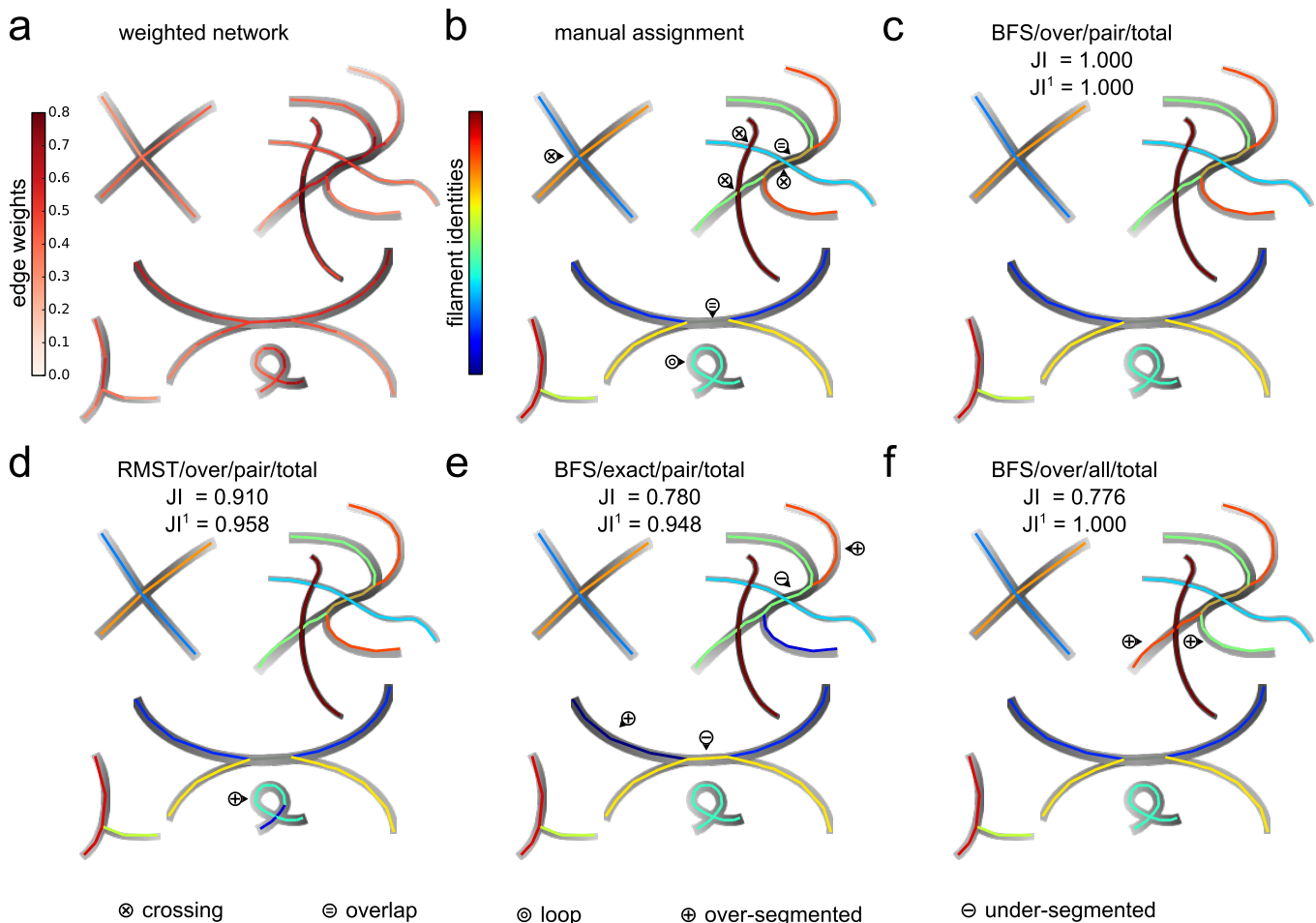
Figure 1. Filament covers of artificial network with crossings, overlaps, and a loop. **(a)** Weighted, artificial network extracted from the underlying drawing, with colour-coded edge weights representing the local image intensity. **(b)** Manual decomposition of the network into filaments with colour-coded indices. The filaments display crossings ($\otimes$), overlaps ($\ominus$), and a loop ($\odot$). **(c)** Filament cover obtained by solving the FCP using the set of input paths generated by a modified breadth-first-search (*BFS*), allowing overlapping filaments (*over*), employing the pairwise roughness measure (*pair*), and by minimising the total roughness of the cover (*total*). The automatically obtained filament cover correctly captures crossings, overlaps, and loops, and agrees excellently with the manual assignment (similarity of the two filament covers is measured by the global Jaccard index, JI, and our modified, structure-aware Jaccard index, JI^1, which reflect the fraction of pairs of all or only adjacent edges that are assigned to the same filament, respectively; here $JI = JI^1 = 1$). The filament identities and colours are matched by solving an assignment problem whereby the total number of edges shared by two filaments, from the manual and automated partitioning, is maximised; the same assignment procedure is used for the remaining panels. **(d)** When using paths obtained from sampling random minimum spanning trees (*RMST*) for the FCP, the closed filament loop is not correctly detected and is over-segmented ($\oplus$). **(e)** When solving the exact FCP (*exact*), the loop is correctly detected. However, overlaps are neglected so that no two filaments share an edge, leading to over- and under-segmentation ($\ominus$). **(f)** When minimising the all-to-all filament roughness (*all*), two half-filaments are interchanged because the maximum weight difference is smaller along the altered filaments.

ured by the structure-aware Jaccard index JI^1. We repeat the procedure for all E edges and then proceed with the removal of E randomly chosen doubles of edges, triplets, up to subsets of 50 edges. As expected, the accuracy of the filament cover typically decreases with the number of removed edges, although removal of some specific edges even leads to an increase in accuracy. However, JI^1 decreases very moderately by less than 0.001 per removed edge on average (cf. Supplemental Material S7). In the second scenario, we assess the robustness of our filament detection approach against noise by adding centred Gaussian noise of increasing standard deviation to the edge weights of the original network. For a given standard deviation, we obtain the optimal filament covers for 100 noisy network instances and compute their similarity, JI^1, to the manual assignment. Again, as expected, the accuracy of the filament cover decreases with increasing noise, but only slightly. On average, increasing the noise by 1% of the original edge weights only decreases JI^1 by less than 0.001. Moreover, we note that with increasing edge

noise the accuracy of the filament cover approaches a constant, non-zero JI^1 which reflects that some information about the filament structure maybe obtained from the topology of the network alone, irrespective of the edge weights (cf. Supplemental Material S7).

Disentangling biological and cosmic filamentous structures

Since we demonstrated the power of the FCP-based approach on contrived filamentous structures, we next proceed with investigating real biological and cosmic filamentous structures (cf. Methods and Supplemental Material S5 for the extraction of the networks; cf. Supplemental Material S9 for an overview of the different stages of our approach). As a first illustrative example of a biological filamentous structure, we extract a weighted network from an image of a hippocampal neuron (Fig. 2a) and manually obtain a filament assignment with several crossings and loops (Fig. 2b, $\otimes$ and $\odot$, respectively). Solving the FCP (same options as in Fig. 1e) yields an automated decomposition which captures well the manual assignment, in particular the two loops (Fig. 2c, $JI^1 = 0.937$). This is further supported by the distributions of filament lengths (measured by the numbers of edges) as well as the distributions of maximal filament angles (measured between adjacent edges), which are statistically indistinguishable between the manual assignment and the automated decomposition (Fig. 2d, black and red; Kolmogorov-Smirnov test p-value $p_{KS} \geq 0.05$). A detailed analysis of the similarity of manual and automated decompositions shows that the classical Rand index RI [52] overestimates the similarity, while the variation of information VI [78] and the Jaccard index JI severely underestimate the similarity between the manual and automated decomposition when compared to the values of the here-proposed RI^1 and JI^1 (Fig. 2e, dotted blue, green, and yellow). The latter measures take into consideration the network structure when comparing two network decompositions (Fig. 2e, solid blue and yellow). We would like to emphasise that the disparities in the estimations of RI and JI result from the consideration of distant, non-adjacent edges which are excluded in RI^1 and JI^1. In addition, we observe that RI^d and JI^d show a non-trivial dependence on the distance, d, between the considered edges, and coincide with the classical similarity measure for large enough distances, i.e., $RI^\infty \equiv RI$ and $JI^\infty \equiv JI$ (cf. Supplemental Material S6 for a detailed discussion).

Finally, different flavours of the FCP may be solved, as mentioned above, to obtain decompositions of varying similarity in comparison to the manual assignment (Fig. 2f). Solving the FCP with paths from the modified BFS, instead of RMSTs, yields consistently higher RI^1- and JI^1-values for the agreement with the manual assignment. This is due to the higher flexibility with respect to the treatment of loops. For the studied networks, a decomposition based on the minimisation of the total roughness yields higher RI^1- and JI^1-values in comparison to the minimisation of the average roughness. In addition, in terms of RI^1 and JI^1, covers allowing for overlaps yield better agreement with the manual assignment, in comparison to those in which each edge is covered by a single path. However, these expected trends are absent or even reversed for the classical similarity measures VI, RI, and JI (cf. Supplemental Material S6), which further justifies the usage of the here-proposed RI^1 and JI^1 for comparing decompositions of networks arising in other network-based analyses (cf. e.g. [82]).

As a second biological example, we investigate the filamentous structure of a plant actin cytoskeleton (Fig. 3a). We create seven manual assignments (one of which is shown in Fig. 3b) for a quantitative comparison with the automated decomposition (Fig. 3c, $JI^1 = 0.655$; same options of the FCP as in Fig. 1e). The agreement of the automated decomposition with the manual assignment is good, despite several over- or under-segmented filaments (Fig. 3c, cf. $\oplus$ and $\ominus$). For a comprehensive assessment of this agreement, we compute the pairwise similarities between the automated and all seven manual filament decompositions (Fig. 3d, upper panel). By comparing the similarities between automated and manual decompositions to the similarities among the different manual decompositions, we find reassuringly that our automated solution is as good as any manual decomposition (Fig. 3d, lower panel, red and black, respectively; cf. independent two-sample Student's t-test p-value $p_t \geq 0.05$). The agreement between the automated decomposition and the reference manual assignment (cf. Fig. 3b) is further confirmed by statistical tests which demonstrate that the two distributions of filament lengths from manual assignment and automated decomposition do not statistically differ (Fig. 3e, upper panel, black and red histograms; cf. $p_{KS} \geq 0.05$). In addition, our results indicate that the filament lengths may be described by a gamma distribution (Fig. 3e, upper panel, dashed lines; maximal likelihood fits of normal, Weibull, and Rayleigh distributions yield higher values for the Akaike information criterion [1]), in agreement with theoretical and experimental studies [24, 33]. Moreover, the distributions of average pairwise filament roughnesses do not differ between manual assignment and automated decomposition (Fig. 3e, lower panel; cf. $p_{KS} \geq 0.05$). We note that the sum of filament roughnesses, R, is larger for the manual assignment of filaments than in the automated decomposition, as expected, as R is the objective function of the minimisation in the FCP-based formulation.

By investigating the relationship between filament length and pairwise roughness, we can distinguish three regions (Fig. 3f): Long filaments typically correspond to actin bundles and exhibit small roughnesses (Fig. $3f_1$), the majority of filaments is shorter with comparable roughnesses (Fig. $3f_2$), and some typically short filaments consist of only one edge with roughness given by the edge weight itself (Fig. $3f_3$; cf. Eq. 1). The angular distribution of filaments indicates that the majority of fila-
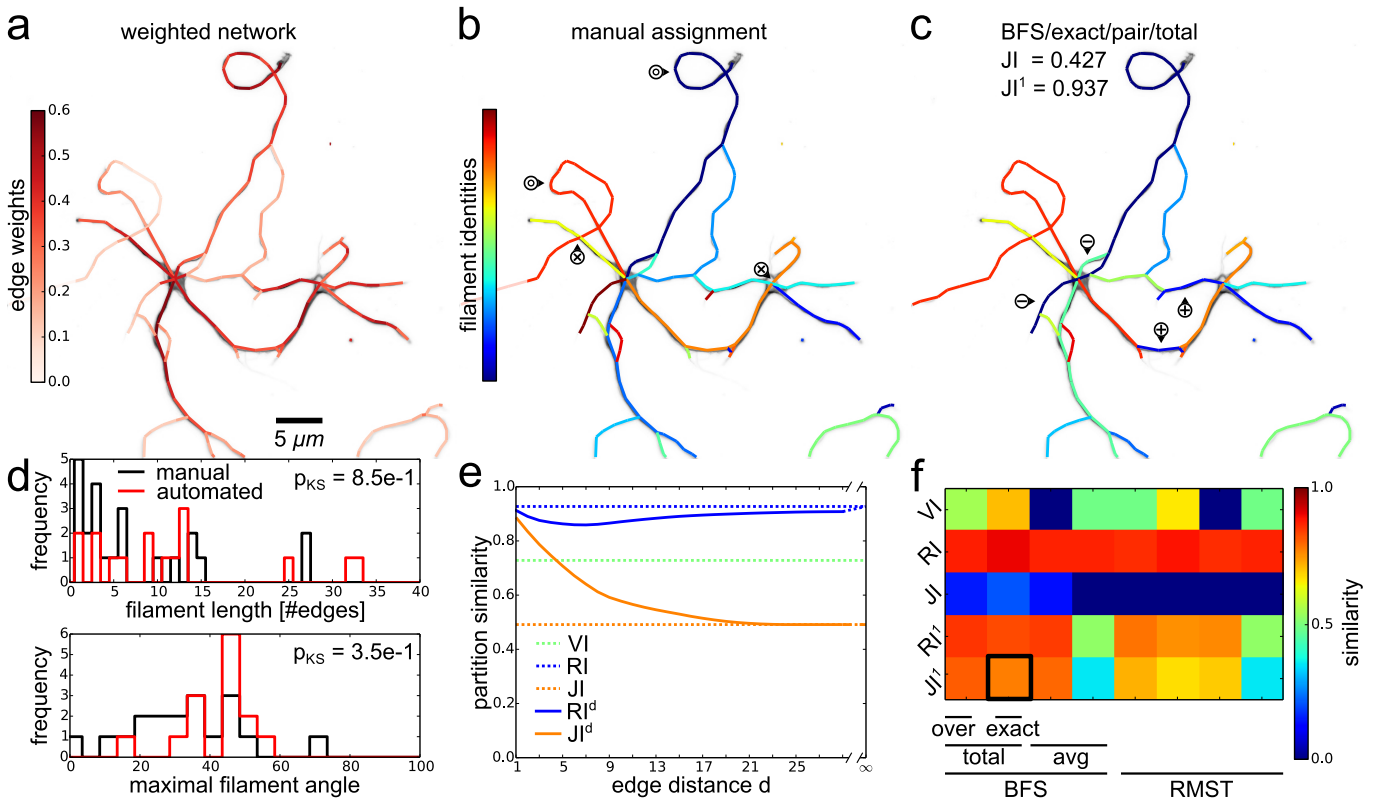
Figure 2. **Filament covers and analyses of neuronal network.** The weighted hippocampal neuronal network is automatically decomposed into filaments by solving the exact FCP (*exact*) for paths from a modified breadth-first search (*BFS*) and by minimising the total (*total*) pairwise filament roughness (*pair*). **(a)** Overlay of fluorescence microscopy image of hippocampal neurons and extracted network with colour-coded edge weights. **(b)** Manual decomposition of the neuronal network into filaments with colour-coded indices and crossings ($\otimes$) and loops ($\circledcirc$). **(c)** Automated partitioning of the network obtained by solving the FCP displays good agreement with the manually obtained partitioning (JI^1 close to 1, see panel (e) for details) with marked illustrative sites of over- ($\oplus$) and under-segmentation ($\ominus$). **(d)** Distributions of numbers of edges per filament (upper panel) as well as distributions of maximum filament angles (lower panel) are similar for manual (black) and automated decomposition (red; Kolmogorov-Smirnov test $p_{KS} \geq 0.05$). **(e)** Different measures of similarity of manual and automated decompositions. The variation of information VI (dashed green) indicates moderate similarity but is not well-defined for general, overlapping decompositions. While the classical Jaccard index JI (dashed yellow) is of small value, the proposed structure-aware extension JI^d increases with decreasing d, i.e., when only edges are considered that are separated by at most d nodes (solid yellow). Moreover, while the classical Rand index RI (dashed blue) is of large value, the proposed structure-aware extension RI^d displays a non-monotonic dependence on d (solid blue). **(f)** Heat map of partition similarities for different similarity measures and options of the FCP, cf. Fig. 1 for a demonstration of the different options. The FCP options which yield the partition shown in (c) are marked by a black rectangle.

ments is aligned parallel to the cell axis (Fig. 3f, dashed grey line) which has been suggested to support longitudinal cell growth [92, 115]. While these reports of longitudinal alignment of the actin cytoskeleton were based on manual or qualitative measurements, our approach facilitates fully automated quantification of the alignment of individual filaments. Our findings show that the length of a filament correlates with its average weight (Fig. 3g; Pearson correlation coefficient $c_P > 0$ and p-value $p_P < 0.05$), i.e., thicker actin bundles stretch across the cell while individual thinner actin filaments are more locally confined, as expected [2, 102].

Finally, we study filament convolutedness, given by the ratio of the length of a filament and the largest side of a bounding box enclosing the filament, used as

a measure for the curvedness of a filament [102]. We find that the convolutedness is slightly negatively correlated with the filament length (Fig. 3i, red; $c_{P,conv} < 0$ and $p_{P,conv} \geq 0.05$), in agreement with previous findings in *Arabidopsis thaliana* pollen grain [102] and other plant species [49]. In contrast to the automated approach used here, the existing studies of filament convolutedness required manual segmentation which may be biased by the user. Generally, and more severely, using a bounding rectangle to compute the convolutedness of a filament is biased by the orientation of the filament with respect to the x- and y-axis of the image. Therefore, we use the maximal filament angle as a non-biased measure for the maximal, local curvedness of a filament. By investigating the relation between the maximal filament angle and
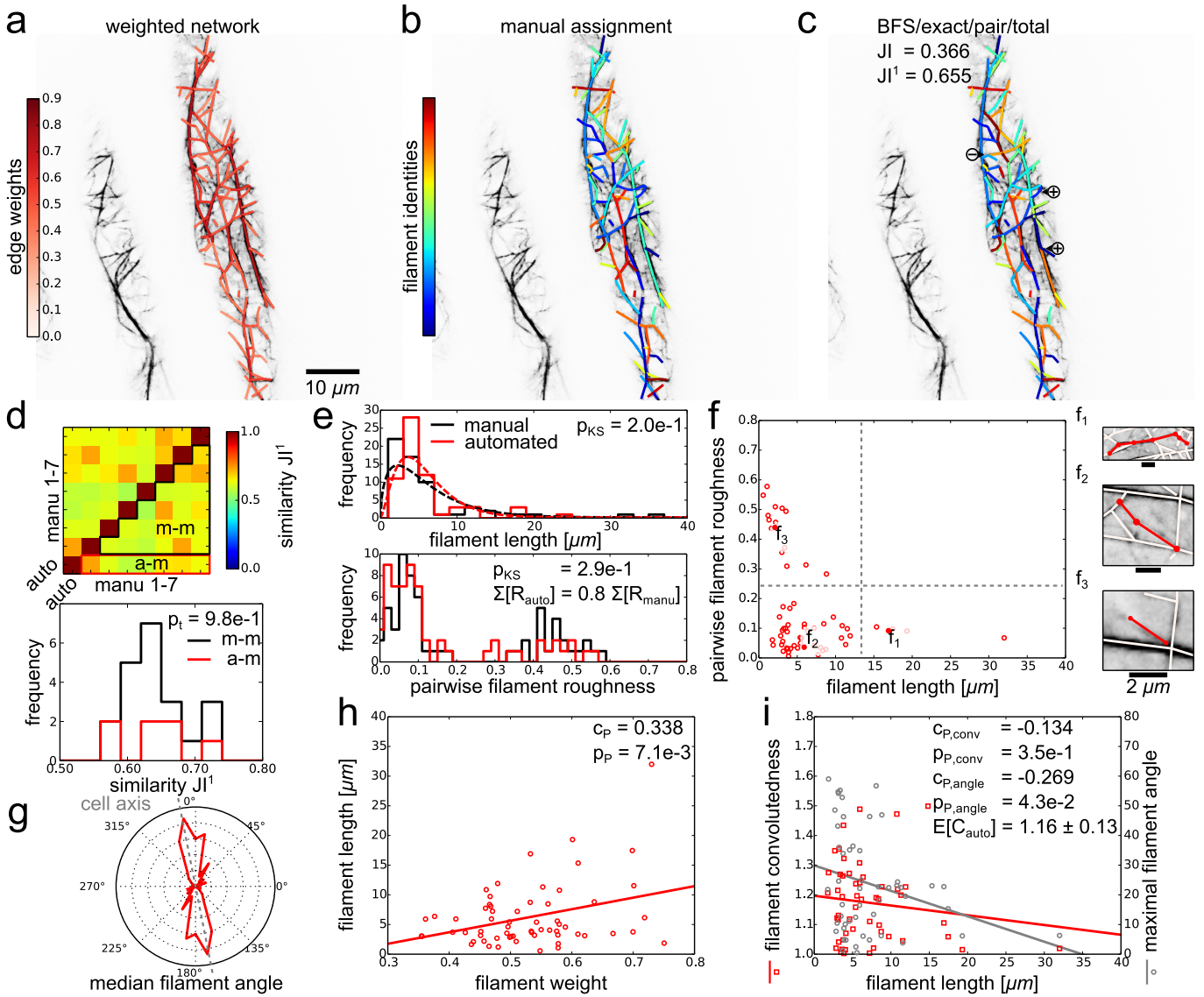
Figure 3. **Filament covers and analyses of cytoskeletal network.** The weighted cytoskeletal network is decomposed automatically by solving the exact FCP (*exact*) for paths from a modified breadth-first search (*BFS*) and by minimising the total (*total*) pairwise filament roughness (*pair*). **(a)** Overlay of confocal microscopy image of an actin cytoskeleton and extracted network with colour-coded edge weights. **(b)** Manual decomposition of the actin cytoskeleton into filaments with colour-coded indices. **(c)** The automated decomposition according to the FCP correctly assigns many of the filaments ($\mathrm{JI}^1 = 0.655$). Some occurrences of over- ($\oplus$) and under-segmentation ($\ominus$) are marked. **(d)** Heat map of similarity between automated (cf. (c)) and seven manual decompositions (cf. e.g. (b); upper panel). The similarities between automated and manual decompositions (red, denoted by a-m) do not differ from similarities among the different manual decompositions (black, m-m; lower panel; cf. independent two-sample Student's t-test p-value $p_t \geq 0.05$). **(e)** Distribution of filament lengths for the manual (black) and automated solution (red) are similar (upper panel; cf. Kolmogorov-Smirnov test p-value $p_{\mathrm{KS}} \geq 0.05$). Maximum likelihood fits of gamma functions are shown as dashed lines. The distributions of pairwise filament roughnesses are similar (lower panel; cf. $p_{\mathrm{KS}} \geq 0.05$), while the total roughness is smaller (cf. summed R-values) for the automated decomposition since it is minimised by the FCP. **(f)** Scatter plot of pairwise filament roughness versus filament length displays three regions, with representative examples $\mathrm{f}_1 - \mathrm{f}_3$ (solid dots): (f_1) For long filaments ($\geq 15\,\mu\mathrm{m}$), the roughness is moderate (< 0.2), as expected for actin bundles; (f_2) The majority of filaments is short ($< 15\,\mu\mathrm{m}$) and of moderate roughness; (f_3) Some typically short filaments show a high roughness (≥ 0.2), namely those which are composed of one network edge only so that their roughness is given by the edge weight itself (cf. Eq. 1). **(g)** The distribution of median filament angles shows that the majority of filaments is aligned parallel to the cell axis (grey dashed line). **(h)** The filament length correlates with the filament weight (cf. linear regression and Pearson correlation coefficient $c_P > 0$ and p-value $p_{\mathrm{P}} < 0.05$) **(i)** Scatter plot of filament convolutedness versus filament length shows a negative but non-significant correlation (cf. red squares, $c_{P,\mathrm{conv}} < 0$, and $p_{P,\mathrm{conv}} \geq 0.05$) with an average convolutedness of $\mathrm{E}\,[C] = 1.16 \pm 0.13$. The maximum filament angle correlates negatively and significantly with the filament length (cf. grey circles, $c_{P,\mathrm{angle}} < 0$, and $p_{P,\mathrm{angle}} < 0.05$), indicating that longer (and thicker, cf. (g)) filaments are less curved.

filament length, we find a significant negative correlation (Fig. 3i, grey; $c_{P,angle} < 0$ and $p_{P,angle} < 0.05$). This negative correlation reflects the known increase in stiffness of actin bundles with increasing bundledness and length [25, 39]. Thus, our approach provides a fast means to investigate this property for individual filaments in a cellular context without laborious manual filament identification.

To further extend these findings, we extract the cytoskeletal networks from 100 frames of a movie of a plant actin cytoskeleton (cf. Methods). For each frame, we compute the optimal filament covers and analyse the filaments. The additional data support our reported findings (Supplemental Material S8).

Moreover, we repeat our analyses of the robustness of our approach against incomplete knowledge of the underlying network structure or noisy edge weights for the cytoskeletal network (cf. discussion of Fig. 1; Supplemental Material S7). In our first scenario, the removal of increasing numbers of edges typically moderately decreases the accuracy of the obtained filament covers, i.e., their agreement with the manual assignment as measured by JI^1. While the removal of some critical edges leads to a more severe decrease in accuracy, there exist edges whose removal leads to an increase in accuracy. On average, the removal of one additional edge decreases JI^1 by around 0.002. Consequently, a loss of 10% of the cytoskeletal network's $E = 179$ edges still yields $JI^1 \approx 0.6$ which is comparable to similarity values between different manual assignments (cf. Fig. 3d; cf. Supplemental Material S7). In our second scenario, the adding of Gaussian noise of increasing standard deviation to the edge weights similarly, as expected, decreases the accuracy of the obtained filament covers. However, this effect is moderate, i.e, increasing the standard deviation by 1% of the original edge weight decreases JI^1 by less than 0.001. Adding noise with a standard deviation of 20% of the original edges weights still yields $JI^1 \approx 0.6$. As for the robustness analyses of the contrived network, for strong noise, JI^1 tends to a constant, non-zero value which suggests that some information about the filament structure may be obtained solely from the network topology, irrespective of the edge weights (cf. discussion of Fig. 1; cf. Supplemental Material S7).

As a final example, we decompose the network of a simulated galaxy cluster (Fig. 4a) into individual galaxy filaments (Fig. 4b). The quantification of galaxy filaments may help to elucidate the acceleration of the universe [99] and improve our understanding of large-scale structure formation [98]. Moreover, studies have revealed gravitational motion of galaxies along individual filaments [6, 35]. Yet, previous studies focused on connected components of the cosmic web, and sought robust methods to identify individual filaments [98, 104]. Our approach confirms the expected discrepancy between the lengths of the components (i.e., the sum of their edge lengths; Fig. 4c, upper panel, grey) and the length of individual filaments (Fig. 4c, upper panel, red; cf. average L-values).

Moreover, the decomposition of the cosmic structures enables analyses of individual filament shapes. For example, the convolutedness which measures the curvedness of a filament shows small values (Fig. 4, lower panel), which are interestingly comparable to those found in the actin cytoskeleton (cf. Fig. 3i; cf. average C-values), indicating the prevalence of straight galaxy filaments.

In Tab. 2, we summarise the quality of the investigated decompositions of different filamentous networks and the options of the underlying FCP (cf. Supplemental Material S8 and S9 for analyses of additional filamentous networks that are not shown in the main text).

DISCUSSION

The decomposition of complex networks into meaningful substructures has facilitated network-based analyses of systems found in nature or designed by humans [80, 95, 100]. These natural and technical networks often embed filaments as basic building units. To enable deeper understanding of network systems with filamentous structure, it is therefore paramount to develop methods for accurate and feasible identification of the underlying filaments. In particular, the distinction between intra- and inter-filament connections enables a more detailed analysis of filamentous structures, including length statistics, spatial alignment, and bending of individual filaments. Such statistics may offer new insights, e.g., into the role of single actin or galaxy filaments in their cellular or cosmic network context, respectively (cf. Figs. 3e-i and 4c).

Here, we proposed a robust optimisation approach to decompose any given weighted network into a set of smooth filaments comprising a filament cover. Since we demonstrated that the filament cover problem is intractable on general networks, we proposed, tested, and validated several alternative approximation schemes. The proposed approximation schemes are gauged at applications from different scientific fields in which filamentous structures naturally arise. We applied our optimisation-based approach on contrived test cases as well as biological and cosmic networks, and showed that it reliably identifies crossing, (non-) overlapping, and looped filaments in agreement with expert-based manual assignments.

Our approach offers a number of advantages over the existing alternatives: (1) The proposed optimisation approach can be applied to any weighted network. In particular, the approach can be readily applied to any network generated from two- or three-dimensional experimental image data typically gathered in biological studies and analyses of man-made systems (e.g. [75, 87, 91, 107]), irrespective of the image source (e.g., light microscopy- or MRI-based). Thus, it may be used to study a variety of natural and technical filamentous structures in search for universal properties which go beyond the characterisation of geometric networks [10].
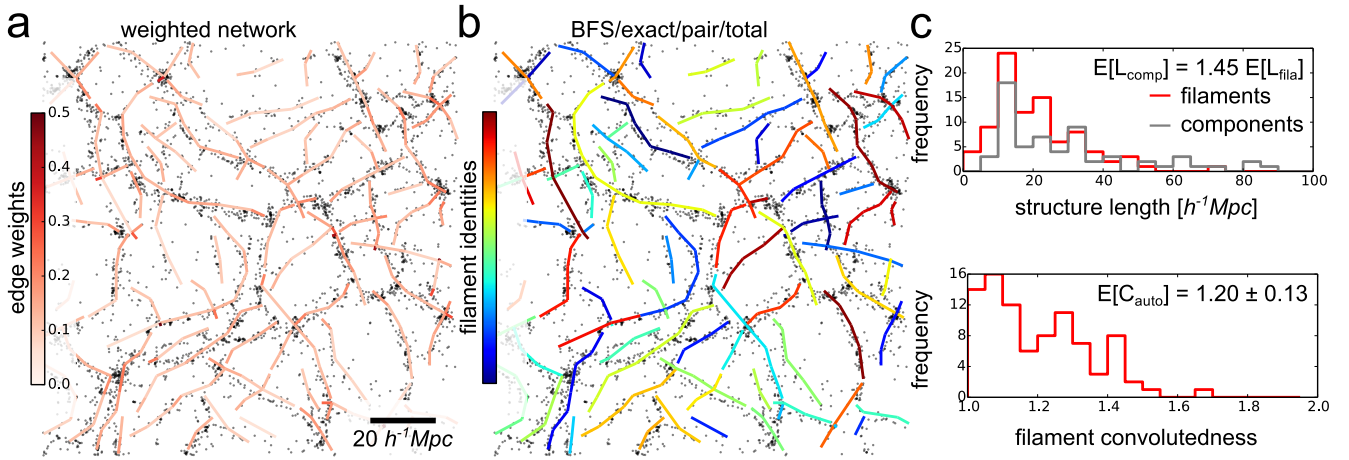
Figure 4. **Filament covers and analyses of cosmic web.** Image data from: Stoica et al., A&A, 434, 423-432, 2005, reproduced with permission © ESO [104]. The cosmic web is decomposed automatically by solving the exact FCP (*exact*) for paths from a modified breadth-first search (*BFS*) and by minimising the total (*total*) pairwise filament roughness (*pair*). Distances are given in h^{-1}Mpc, where currently $h \approx 0.7$ is the dimensionless Hubble parameter [29]. **(a)** Overlay of simulated galaxy clusters and extracted network with colour-coded edge weights. **(b)** Automated decomposition of the cosmic web into galaxy filaments with colour-coded indices. **(c)** The length distribution of galaxy filaments exhibits a peak around $20\,h^{-1}$Mpc and levels off for larger lengths (upper panel, red). As a comparison, the distribution of the total lengths of the connected components levels off more slowly and overestimates the average filament length by a factor of 1.45 (upper panel, grey; cf. average L-values). The distribution of the convolutedness of galaxy filaments suggests a prevalence of straight filaments and its average is comparable to that of the actin network (cf. 3i; cf. $\mathrm{E}\,[C] = 1.20 \pm 0.13$).

		Figure	**Options**	**Similarity**				
				VI	RI $(\equiv \mathrm{RI}^\infty)$	JI $(\equiv \mathrm{JI}^\infty)$	RI^1	JI^1
artificial	overlaps + loop	1	*BFS over pair tot*	0.792	1.000	1.000	1.000	1.000
	grid-like	S5b	*BFS exact pair tot*	0.889	0.962	0.742	0.941	0.872
neural	hippocampus	2	*BFS exact pair tot*	0.848	0.906	0.427	0.954	0.937
	retina	S5d	*BFS exact pair tot*	0.792	0.963	0.397	0.905	0.883
cytoskeletal	actin (FABD-labelled)	3	*BFS exact pair tot*	0.829	0.976	0.366	0.854	0.655
	actin (Lifeact-labelled)	S5f	*BFS exact pair tot*	0.530	0.929	0.193	0.838	0.701
cosmic	galaxy cluster (sparse)	4	*BFS exact pair tot*	no manual assignment				
	galaxy cluster (dense)	S5h	*BFS exact pair tot*	for comparison				

Table 2. **Quality of filament covers of artificial, biological, and cosmic networks in comparison to manual decompositions.** A given network is decomposed into filaments by solving the FCP with different options: The initial set of paths is obtained from a modified breadth-first search (*BFS*) or sampling of random minimum spanning trees (*RMST*), the filaments may overlap (*over*) or not (*exact*), a pairwise (*pair*) or all-to-all filament roughness measure (*all*) is used, and the total (*total*) or average (*avg*) roughness is minimised. The table displays the investigated filament covers with high similarity to the manual assignments.

(2) Our approach facilitates the establishment of a link between the dynamics of individual filaments and the dynamics of the whole network. While the dynamics of individual filaments is guided by typically molecular, local processes, the behaviour of the entire filamentous structure incorporates and responds to stimuli across different scales. Therefore, the proposed approach provides the starting point towards network-oriented analysis of filaments. More specifically, the filament covers may even be used to track mobile filaments, as has been proposed for images of a few filaments using open contours [97], providing a venue for fruitful applications of the method.

(3) The different options of our approach, e.g., different measures of the filament roughness, enable flexible and intuitive customisation for different types of networks. For example, the filament roughness measure may include a penalty for filament bending in networks of straight, stiff filaments (such as microtubules [43, 111]), or a penalty for length deviations in networks of filaments of mostly uniform length (such as synthetic polymers that are used, e.g., in drug delivery systems [3, 48]).

(4) At the same time, our approach to disentangle a given network is parsimonious, i.e., it has a strictly limited number of categorical options which allow testing of all possible combinations ($4^2 = 16$ in total). In contrast, approaches which rely on multiple continuous parameters require data-specific and computationally expensive gauging of the parameters [65, 89]. When compared to

approaches which detect filaments directly from image data, however, the parsimony of our approach is counterbalanced by the parameter requirements of the preceding network extraction procedure.

(5) Nevertheless, approaches that detect filaments directly from image data typically rely on local optimisation schemes and thus, e.g., on the order of filament initialisations and definitions of local filament properties [76, 77, 88, 121]. In contrast, our approach offer the advantage that the decomposition into filaments is performed in a single optimisation step which holistically considers the global structure of both filaments and network.

(6) Finally, since our approach replies on a general network representation, it may be applied also to networks obtained from other, e.g., open contour-based methods which often do not capture filament overlaps and result in fragmented filaments [120, 121]. In a post-processing step, these fragments may be conveniently merged using our network-based approach (cf. Supplemental Material 10).

Yet, some caution is warranted: (1) The available options of the FCP yield different decompositions. We showed that paths sampled from a modified BFS enable more flexible and more accurate decompositions in comparison to paths sampled from RMSTs (cf. Fig. 1); in contrast to minimising the the average roughness, the minimisation of the total roughness favours longer filaments in better accordance with the manual assignments (cf. Fig. 1); moreover, since filament overlaps in biological systems may lead to an abrupt increase in apparent filament thickness, the proposed all-to-all filament roughness may be more suitable to study such situations than the pairwise filament roughness which favours filaments of slowly varying thickness. Therefore, the suitable choice of feasible and suitable options has to be further investigated. For example, for the actin cytoskeletal networks, it is not obvious if overlapping filaments should be preferred over non-overlapping filaments and if the pairwise roughness is a better measure of filament quality than the all-to-all roughness. Yet, such decision problems are innate not only to all automated decomposition algorithms, but also to the manual assignment based on which the performance is assessed. Thus, exploring different decomposition options by an expert in the field may hint at the right choice.

(2) The quality of the filament cover clearly depends on the quality of the input network. To this end, several algorithms have been proposed for the extraction of various types of networks from image data with low error rates [11, 27, 77, 83, 89, 121]. Moreover, we investigated different scenarios to test the robustness of our approach against incomplete knowledge of the underlying network structure as well as low signal-to-noise ratios and found that the accuracy of the filament cover is only moderately affected by these obstacles (cf. Supplemental Material S7).

(3) Another issue are the computational requirements of the FCP. Although our proposed approximation scheme employs a modified BFS and a binary linear program which run fast on the tested networks, it may become infeasible for larger networks comprising more edges or nodes of larger degrees. Therefore, future efforts may focus on devising algorithms which approximate the FCP by employing local searches, i.e., without sampling a large number of paths for the proposed set cover-based approximation scheme.

(4) Finally, we note that many polymers are not simple linear chains but branched tree-like structures [53, 108]. Also many neurons may be naturally described as tree-like structures [5, 114]. Our approach can be extended to account for these cases, thus, opening a new field of research. To this end, covering networks with more complex structures, such as stars [28, 67, 106] or, more generally, trees [34, 51] may be employed. Due to intractability of these problems, investigation of approximation schemes like our set cover formulation will be needed. A central question will be the development of measures for the quality of a given star or tree cover.

In conclusion, by decomposing technically and biologically relevant filamentous structures into their constitutive filaments, our approach allows to see both the wood and the trees.

AUTHOR CONTRIBUTIONS

D.B. implemented the method and analysed the data; D.B. and Z.N. developed the method, showed the computational complexity of the problem, and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

[1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans Autom Control*, 19(6):716–723.

[2] Akkerman, M., Overdijk, E. J., Schel, J. H., Emons, A. M. C., and Ketelaar, T. (2011). Golgi body motility in the plant cell cortex correlates with actin cytoskeleton

organization. *Plant Cell Physiol*, 52(10):1844–1855.

[3] Ali, M. and Brocchini, S. (2006). Synthetic approaches to uniform polymers. *Adv Drug Delivery Rev*, 58(15):1671–1687.

[4] Andreatta, G. and Mason, F. (1995). Path covering problems and testing of printed circuits. *Discrete Appl Math*, 62(1):5–13.

[5] Ascoli, G. A., Donohue, D. E., and Halavi, M. (2007). Neuromorpho.org: a central resource for neuronal morphologies. *J Neurosci*, 27(35):9247–9251.

[6] Aubert, D., Pichon, C., and Colombi, S. (2004). The origin and implications of dark matter anisotropic cosmic infall on L haloes. *Mon Not R Astron Soc*, 352(2):376–398.

[7] Bae, E., Bailey, J., and Dong, G. (2010). A clustering comparison measure using density profiles and its application to the discovery of alternate clusterings. *Data Min Knowl Disc*, 21(3):427–471.

[8] Bai, M., Missel, A. R., Levine, A. J., and Klug, W. S. (2011). On the role of the filament length distribution in the mechanics of semiflexible networks. *Acta Biomater*, 7(5):2109–2118.

[9] Bálint, S., Verdeny Vilanova, I., Sandoval Álvarez, A., and Lakadamyali, M. (2013). Correlative live-cell and superresolution microscopy reveals cargo transport dynamics at microtubule intersections. *Proc Natl Acad Sci*, 110(9):3375–3380.

[10] Barthélemy, M. (2011). Spatial networks. *Phys Rep*, 499(1):1–101.

[11] Baumgarten, W. and Hauser, M. J. (2012). Computational algorithms for extraction and analysis of two-dimensional transportation networks. *J Comput Interdiscip Sci*, 3:107–16.

[12] Bausch, A. and Kroy, K. (2006). A bottom-up approach to cell mechanics. *Nat Phys*, 2(4):231–238.

[13] Bennett, M. (1977). *Handbook of Physiology*, chapter Electrical transmission: a functional analysis and comparison to chemical transmission. Wiley Online Library, Bethesda, 1st edition.

[14] Bertossi, A. A. (1981). The edge Hamiltonian path problem is NP-complete. *Inform Process Lett*, 13(4):157–159.

[15] Biswas, S., Durocher, S., Mondal, D., and Nishat, R. I. (2012). Hamiltonian paths and cycles in planar graphs. In *Combinatorial Optimization and Applications*, pages 83–94. Springer, Berlin.

[16] Bond, J., Kofman, L., and Pogosyan, D. (1996). How filaments of galaxies are woven into the cosmic web. *Nature*, 380(6575):603–606.

[17] Bond, N. A., Strauss, M. A., and Cen, R. (2010). Crawling the cosmic network: identifying and quantifying filamentary structure. *Mon Not R Astron Soc*, 409(1):156–168.

[18] Boudaoud, A., Burian, A., Borowska-Wykret, D., Uyttewaal, M., Wrzalik, R., Kwiatkowska, D., and Hamant, O. (2014). FibrilTool, an ImageJ plug-in to quantify fibrillar structures in raw microscopy images. *Nat Protoc*, 9(2):457–463.

[19] Braitenberg, V. and Schüz, A. (1998). *Cortex: Statistics and Geometry of Neuronal Connectivity*. Springer, Berlin, 1st edition.

[20] Brandner, D. and Withers, G. (2014). Development of the axon and dendritic arbors in cultured hippocampal neurons. CC-BY 3.0, http://www.cellimagelibrary.org/contributors/742096, Date of access: 07/09/2015.

[21] Brešar, B., Kardoš, F., Katrenič, J., and Semanišin, G. (2011). Minimum k-path vertex cover. *Discrete Appl Math*, 159(12):1189–1195.

[22] Breuer, D., Ivakov, A., Sampathkumar, A., Hollandt, F., Persson, S., and Nikoloski, Z. (2014). Quantitative analyses of the plant cytoskeleton reveal underlying organizational principles. *J R Soc Interface*, 11(97):20140362.

[23] Buchin, K., Knauer, C., Kriegel, K., Schulz, A., and Seidel, R. (2007). On the number of cycles in planar graphs. In *Computing and Combinatorics*, pages 97–107. Springer, Berlin.

[24] Burlacu, S., Janmey, P., and Borejdo, J. (1992). Distribution of actin filament lengths measured by fluorescence microscopy. *Am J Physiol-Cell Ph*, 262(3):C569–C577.

[25] Claessens, M. M., Bathe, M., Frey, E., and Bausch, A. R. (2006). Actin-binding proteins sensitively mediate F-actin bundle stiffness. *Nat Mater*, 5(9):748–753.

[26] Coen, M. H., Ansari, M. H., and Fillmore, N. (2010). Comparing clusterings in space. In *Proc 27th Internat Conf Mach Learn*, pages 231–238, Madison. Omnipress.

[27] Cohen, A. R., Roysam, B., and Turner, J. N. (1994). Automated tracing and volume measurements of neurons from 3D confocal fluorescence microscopy data. *J Microsc*, 173(2):103–114.

[28] Cohen, E. and Tarsi, M. (1991). NP-completeness of graph decomposition problems. *J Complexity*, 7(2):200–212.

[29] Croton, D. J. (2013). Damn you, little h! (or, real-world applications of the hubble constant using observed and simulated data). *Publ Astron Soc Aust*, 30:e052.

[30] Dahl, J. and Vandenberghe, L. (2006). Cvxopt: a Python package for convex optimization. http://abel.ee.ucla.edu/cvxopt/, Date of access: 07/09/2015.

[31] Denœud, L. and Guénoche, A. (2006). Comparison of distance indices between partitions. In *Data Science and Classification*, pages 21–28. Springer, Berlin.

[32] Eccles, J. C. (1982). The synapse: from electrical to chemical transmission. *Annu Rev Neurosci*, 5(1):325–339.

[33] Ermentrout, G. B. and Edelstein-Keshet, L. (1998). Models for the length distributions of actin filaments: II. polymerization and fragmentation by gelsolin acting together. *B Math Biol*, 60(3):477–503.

[34] Even, G., Garg, N., Könemann, J., Ravi, R., and Sinha, A. (2004). Min–max tree covers of graphs. *Op Res Lett*, 32(4):309–315.

[35] Faltenbacher, A., Gottlöber, S., Kerscher, M., and Mueller, V. (2002). Correlations in the orientations of galaxy clusters. *Astron Astrophys*, 395(1):1–9.

[36] Finlay, J. (2005). PyGTK 2.0 Tutorial. http://www.pygtk.org/dist/pygtk2-tut.pdf, Date of access: 07/09/2015.

[37] Frangi, A. F., Niessen, W. J., Vincken, K. L., and Viergever, M. A. (1998). Multiscale vessel enhancement filtering. In *Med Image Comput Comput-Ass Interv*, pages 130–137. Springer, Berlin.

[38] Galkovskyi, T., Mileyko, Y., Bucksch, A., Moore, B., Symonova, O., Price, C. A., Topp, C. N., Iyer-Pascuzzi, A. S., Zurek, P. R., Fang, S., et al. (2012). GiA roots:

software for the high throughput analysis of plant root system architecture. *BMC Plant Biol*, 12(1):116.

[39] Gardel, M. L., Shin, J. H., MacKintosh, F. C., Mahadevan, L., Matsudaira, P., and Weitz, D. A. (2004). Elastic behavior of cross-linked and bundled actin networks. *Science*, 304(5675):1301–1305.

[40] Garey, M. R. and Johnson, D. S. (1979). *Computers and intractability: a guide to NP-completeness*. WH Freeman, New York, 1st edition.

[41] Garey, M. R., Johnson, D. S., and Tarjan, R. E. (1976). The planar Hamiltonian circuit problem is NP-complete. *SIAM J Comput*, 5(4):704–714.

[42] Gibson, S. and Fan, Y. (2006). Coronal prominence structure and dynamics: a magnetic flux rope interpretation. *J Geophys Res-Space*, 111(A12).

[43] Gittes, F., Mickey, B., Nettleton, J., and Howard, J. (1993). Flexural rigidity of microtubules and actin filaments measured from thermal fluctuations in shape. *J Cell Biol*, 120(4):923–934.

[44] Goldberg, M. K., Hayvanovych, M., and Magdon-Ismail, M. (2010). Measuring similarity between sets of overlapping clusters. In *IEEE 2nd Internat Conf Social Comput*, pages 303–308, Minneapolis. IEEE.

[45] Griva, I., Nash, S. G., and Sofer, A. (2009). *Linear and nonlinear optimization*. SIAM, Philadelphia, 2nd edition.

[46] Hagberg, A., Swart, P., and S Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX. In *Proc 7th Python Sci Conf*, Pasadena. LANL.

[47] Haralick, R. M., Sternberg, S. R., and Zhuang, X. (1987). Image analysis using mathematical morphology. *IEEE Trans Pattern Anal Mach Intell*, 4:532–550.

[48] Hartmann, L. and Börner, H. (2009). Precision polymers: monodisperse, monomer-sequence-defined segments to target future demands of polymers in medicine. *Adv Mater*, 21(32-33):3425–3431.

[49] Henty-Ridilla, J. L., Li, J., Blanchoin, L., and Staiger, C. J. (2013). Actin dynamics in the cortical array of plant cells. *Curr Opin Plant Biol*, 16(6):678–687.

[50] Himsolt, M. (1997). GML: a portable graph file format. http://www.fmi.uni-passau.de/graphlet/gml/gml-tr.html, Date of access: 07/09/2015.

[51] Horak, P. and McAvaney, K. (2008). On covering vertices of a graph by trees. *Discrete Math*, 308(19):4414–4418.

[52] Hubert, L. and Arabie, P. (1985). Comparing partitions. *J Classif*, 2(1):193–218.

[53] Inoue, K. (2000). Functional dendrimers, hyperbranched and star polymers. *Prog Polym Sci*, 25(4):453–571.

[54] Jacques, E., Buytaert, J., Wells, D. M., Lewandowski, M., Bennett, M. J., Dirckx, J., Verbelen, J.-P., and Vissenberg, K. (2013). Microfilament Analyzer, an image analysis tool for quantifying fibrillar orientation, reveals changes in microtubule organization during gravitropism. *Plant J*, 74(6):1045–1058.

[55] Kandel, E., Schwartz, J., and Jessell, T. (2000). *Principles of Neural Science*. McGraw-Hill, New York, 4 edition.

[56] Karp, R. M. (1972). *Reducibility among combinatorial problems*. Springer, New York, 1st edition.

[57] Kasza, K. E., Broedersz, C. P., Koenderink, G. H., Lin, Y. C., Messner, W., Millman, E. A., Nakamura, F., Stossel, T. P., Mackintosh, F. C., and Weitz, D. A. (2010). Actin filament length tunes elasticity of flexibly cross-linked actin networks. *Biophys J*, 99(4):1091–1100.

[58] Katifori, E. and Magnasco, M. O. (2012). Quantifying loopy network architectures. *PLoS One*, 7(6):e37994.

[59] Klemm, D., Heublein, B., Fink, H.-P., and Bohn, A. (2005). Cellulose: fascinating biopolymer and sustainable raw material. *Angew Chem Int Ed*, 44(22):3358–3393.

[60] Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Nav Res Logist Q*, 2(1):83–97.

[61] Kumar, S., Maxwell, I. Z., Heisterkamp, A., Polte, T. R., Lele, T. P., Salanga, M., Mazur, E., and Ingber, D. E. (2006). Viscoelastic retraction of single living stress fibers and its impact on cell shape, cytoskeletal organization, and extracellular matrix mechanics. *Biophys J*, 90(10):3762–3773.

[62] Lancichinetti, A. and Fortunato, S. (2009a). Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Phys Rev E*, 80(1):016118.

[63] Lancichinetti, A. and Fortunato, S. (2009b). Community detection algorithms: a comparative analysis. *Phys Rev E*, 80(5):056117.

[64] Lancichinetti, A., Fortunato, S., and Kertész, J. (2009). Detecting the overlapping and hierarchical community structure in complex networks. *New J Phys*, 11(3):033015.

[65] Leandro, J. J. G., Cesar-Jr, R. M., and Costa, L. d. (2009). Automatic contour extraction from 2D neuron images. *J Neurosci Methods*, 177(2):497–509.

[66] Lichtman, J., Livet, J., and Sanes, J. (2008). A technicolour approach to the connectome. *Nat Rev Neurosci*, 9:417–422.

[67] Lin, C. and Shyu, T.-W. (1996). A necessary and sufficient condition for the star decomposition of complete graphs. *J Graph Theor*, 23(4):361–364.

[68] Lin, G., Cai, Z., and Lin, D. (2006). Vertex covering by paths on trees with its applications in machine translation. *Inform Process Lett*, 97(2):73–81.

[69] Lin, R., Olariu, S., and Pruesse, G. (1995). An optimal path cover algorithm for cographs. *Comput Math Appl*, 30(8):75–83.

[70] Linderoth, J. T. and Ralphs, T. K. (2005). Noncommercial software for mixed-integer linear programming. *Int Prog Theor Pract*, 3:253–303.

[71] Liu, B. (2010). *The plant cytoskeleton*. Springer, New York, 2nd edition.

[72] Lobet, G., Pound, M. P., Diener, J., Pradal, C., Draye, X., Godin, C., Javaux, M., Leitner, D., Meunier, F., Nacry, P., et al. (2015). Root System Markup Language: toward a unified root architecture description language. *Plant Physiol*, 167(3):617–627.

[73] Lu, L., Oswald, S. J., Ngu, H., and Yin, F. C.-P. (2008). Mechanical properties of actin stress fibers in living cells. *Biophys J*, 95(12):6060–6071.

[74] Mackay, D., Karpen, J., Ballester, J., Schmieder, B., and Aulanier, G. (2010). Physics of solar prominences: II. magnetic structure and dynamics. *Space Sci Rev*, 151(4):333–399.

[75] Masland, R. H. (2001). The fundamental plan of the retina. *Nat Neurosci*, 4(9):877–886.

[76] Mayerich, D. M. and Keyser, J. (2008). Filament tracking and encoding for complex biological networks. In *Proc 2008 ACM Symp Solid Phys Model*, pages 353–358, New York. ACM.

[77] Meijering, E. (2010). Neuron tracing in perspective. *Cytometry A*, 77(7):693–704.

[78] Meilă, M. (2003). Comparing clusterings by the variation of information. In *Learning Theory and Kernel Machines*, pages 173–187. Springer, Berlin.

[79] Meilă, M. (2005). Comparing clusterings: an axiomatic view. In *Proc 22nd Internat Conf Mach Learn*, pages 577–584, New York. ACM.

[80] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, 298:824–827.

[81] Moon, R. J., Martini, A., Nairn, J., Simonsen, J., and Youngblood, J. (2011). Cellulose nanomaterials review: structure, properties and nanocomposites. *Chem Soc Rev*, 40(7):3941–3994.

[82] Newman, M. (2012). Communities, modules and large-scale structure in networks. *Nat Phys*, 8(1):25–31.

[83] Obara, B., Grau, V., and Fricker, M. D. (2012). A bioimage informatics approach to automatically extract complex fungal networks. *Bioinformatics*, 28(18):2374–2381.

[84] Olivier, B., Rohwer, J., and Hofmeyr, J.-H. (2002). Modelling cellular processes with Python and Scipy. *Mol Biol Rep*, 29(1):249–254.

[85] Osunbayo, O., Butterfield, J., Bergman, J., Mershon, L., Rodionov, V., and Vershinin, M. (2015). Cargo transport at microtubule crossings: evidence for prolonged tug-of-war between kinesin motors. *Biophys J*, 108(6):1480–1483.

[86] Pak-Ken, W. (1999). Optimal path cover problem on block graphs. *Lect Notes Comput Sc*, 225(1):163–169.

[87] Paredez, A. R., Somerville, C. R., and Ehrhardt, D. W. (2006). Visualization of cellulose synthase demonstrates functional association with microtubules. *Science*, 312(5779):1491–1495.

[88] Peng, H., Hawrylycz, M., Roskams, J., Hill, S., Spruston, N., Meijering, E., and Ascoli, G. A. (2015). BigNeuron: large-scale 3D neuron reconstruction from optical microscopy images. *Neuron*, 87(2):252–256.

[89] Qiu, J. and Li, F.-F. (2014). Quantitative morphological analysis of curvilinear network for microscopic image based on individual fibre segmentation. *J Microsc*, 256(3):153–165.

[90] Rao, A. S. and C., P. R. (1990). Linear algorithm for optimal path cover problem on interval graphs. *Inform Process Lett*, 35(3):149–153.

[91] Riedl, J., Crevenna, A. H., Kessenbrock, K., Yu, J. H., Neukirchen, D., Bista, M., Bradke, F., Jenne, D., Holak, T. A., Werb, Z., et al. (2008). Lifeact: a versatile marker to visualize F-actin. *Nat Methods*, 5(7):605–607.

[92] Sampathkumar, A., Lindeboom, J. J., Debolt, S., Gutierrez, R., Ehrhardt, D. W., Ketelaar, T., and Persson, S. (2011). Live cell imaging reveals structural associations between the actin and microtubule cytoskeleton in Arabidopsis. *Plant Cell*, 23(6):2302–2313.

[93] Saporta, G. and Youness, G. (2002). Comparing two partitions: some proposals and experiments. In *Compstat*, pages 243–248. Springer, Berlin.

[94] Schrijver, A. (1998). *Theory of Linear and Integer Programming*. Wiley, New York, 1st edition.

[95] Shen-Orr, S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. *Nat Genet*, 31(1):64–68.

[96] Shih, Y.-L. and Rothfield, L. (2006). The bacterial cytoskeleton. *Microbiol Mol Biol Rev*, 70(3):729–754.

[97] Smith, M. B., Li, H., Shen, T., Huang, X., Yusuf, E., and Vavylonis, D. (2010). Segmentation and tracking of cytoskeletal filaments using open active contours. *Cytoskeleton*, 67(11):693–705.

[98] Sousbie, T., Pichon, C., Colombi, S., Novikov, D., and Pogosyan, D. (2008a). The 3D skeleton: tracing the filamentary structure of the universe. *Mon Not R Astron Soc*, 383(4):1655–1670.

[99] Sousbie, T., Pichon, C., Courtois, H., Colombi, S., and Novikov, D. (2008b). The three-dimensional skeleton of the SDSS. *Astrophys J Lett*, 672(1):L1.

[100] Sporns, O. and Kötter, R. (2004). Motifs in brain networks. *PLoS Biol*, 2:e369.

[101] Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput Biol*, 1(4):e42.

[102] Staiger, C. J., Sheahan, M. B., Khurana, P., Wang, X., McCurdy, D. W., and Blanchoin, L. (2009). Actin filament dynamics are dominated by rapid growth and severing activity in the Arabidopsis cortical array. *J Cell Biol*, 184(2):269–280.

[103] Stamm, A. J. (1964). *Wood and cellulose science*. Ronald Press Co, New York, 1st edition.

[104] Stoica, R. S., Martinez, V. J., Mateu, J., and Saar, E. (2005). Detection of cosmic filaments using the Candy model. *Astron Astrophys*, 434(2):423–432.

[105] Supowit, K. J. (1983). The relative neighborhood graph, with an application to minimum spanning trees. *J ACM*, 30(3):428–448.

[106] Tarsi, M. (1981). On the decomposition of a graph into stars. *Discrete Math*, 36(3):299–304.

[107] Tero, A., Takagi, S., Saigusa, T., Ito, K., Bebber, D. P., Fricker, M. D., Yumiki, K., Kobayashi, R., and Nakagaki, T. (2010). Rules for biologically inspired adaptive network design. *Science*, 327(5964):439–442.

[108] Tomalia, D. A. and Frechet, J. M. (2001). *Dendrimers and other dendritic polymers*. Wiley, New York, 1st edition.

[109] Toussaint, G. T. (1980). The relative neighbourhood graph of a finite planar set. *Lect Notes Comput Sci*, 12(4):261–268.

[110] Tully, R. B., Courtois, H., Hoffman, Y., and Pomarède, D. (2014). The Laniakea supercluster of galaxies. *Nature*, 513(7516):71–73.

[111] van Mameren, J., Vermeulen, K. C., Gittes, F., and Schmidt, C. F. (2009). Leveraging single protein polymers to measure flexural rigidity. *J Phys Chem B*, 113(12):3837–3844.

[112] Van Rossum, G. and Drake, F. L. (2011). *Python Language Reference Manual*. Network Theory Ltd, Godalming, 1st edition.

[113] Vazirani, V. V. (2001). *Approximation algorithms*. Springer, Berlin, 1st edition.

[114] Verwer, R. W. and van Pelt, J. (1983). A new method for the topological analysis of neuronal tree structures. *J Neurosci Meth*, 8(4):335–351.

[115] Waller, F. and Nick, P. (1997). Response of actin microfilaments during phytochrome-controlled growth of maize seedlings. *Protoplasma*, 200(3-4):154–162.

[116] Wickstead, B. and Gull, K. (2011). The evolution of the cytoskeleton. *J Cell Biol*, 194(4):513–525.

[117] Wolsey, L. A. and Nemhauser, G. L. (1999). *Integer and Combinatorial Optimization*. Wiley-Interscience, New York, 1st edition.

[118] Wood, S. T., Dean, B. C., and Dean, D. (2013). A linear programming approach to reconstructing subcellular structures from confocal images for automated generation of representative 3D cellular models. *Med Image Anal*, 17(3):337–347.

[119] Wu, T.-H. (1997). A note on a global approach for general 0–1 fractional programming. *Eur J Oper Res*, 101(1):220–223.

[120] Xu, T., Vavylonis, D., and Huang, X. (2014). 3D actin network centerline extraction with multiple active contours. *Med Image Anal*, 18(2):272–284.

[121] Xu, T., Vavylonis, D., Tsai, F.-C., Koenderink, G. H., Nie, W., Yusuf, E., I-Ju Lee, Wu, J.-Q., and Huang, X. (2015). SOAX: a software for quantification of 3D biopolymer networks. *Sci Rep*, 5:9081.

[122] Yue, D., Guillén-Gosálbez, G., and You, F. (2013). Global optimization of large-scale mixed-integer linear fractional programming problems: a reformulation-linearization method and process scheduling applications. *AIChE J*, 59(11):4255–4272.

[123] Zhou, D., Li, J., and Zha, H. (2005). A new mallows distance based metric for comparing clusterings. In *Proc 22nd Internat Conf Mach Learn*, pages 1028–1035, New York. ACM.

[124] Zhu, J., Ingram, P. A., Benfey, P. N., and Elich, T. (2011). From lab to field, new approaches to phenotyping root system architecture. *Curr Opin Plant Biol*, 14(3):310–317.

DeFiNe: an optimisation-based method for robust disentangling of filamentous networks

Supplemental Material S1-S10

David Breuer[1,*] and Zoran Nikoloski[1]

[1]*Systems Biology and Mathematical Modeling, Max Planck Institute of Molecular Plant Physiology, Am Muehlenberg 1, 14476 Potsdam, Germany*

[*]*breuer@mpimp-golm.mpg.de*

SUPPLEMENTAL MATERIAL S1: MATHEMATICAL FORMULATION OF THE FILAMENT COVER PROBLEM

The structure of a filamentous network is described by a weighted geometric graph $G = (\mathcal{N}, \mathcal{E})$ with $N = |\mathcal{N}|$ nodes and $E = |\mathcal{E}|$ undirected, weighted edges. Edges represent filament segments and nodes represent their endpoints. The positions of the nodes are v_n, $n \in \mathcal{N}$, whereby, typically, $v_n \in \mathbb{R}^2$ or $v_n \in \mathbb{R}^3$ for networks extracted from image data. We focus on geometric networks because filaments are embedded in space, but our approach is readily applicable to non-geometric graphs. The edge weights are w_e, $e := (n_0, n_1) \in \mathcal{E}$ and $n_0, n_1 \in \mathcal{N}$.

To decompose the graph G into individual filaments it is natural to decompose it into paths, i.e., to solve a path cover problem (PCP). The PCP has been intensively studied on different types of graphs and with various restrictions (e.g. [4, 21, 68, 69, 86, 90]). There are several potential routes (cf. [4] for an overview of the PCP for testing printed circuits): (1) We may either use node- or edge-paths, where a path $p = (a_{p,1}, \ldots, a_{p,P})$ is an ordered sequence of $P = |p|$ pairwise adjacent nodes ($a \in \mathcal{N}$) or edges ($a \in \mathcal{E}$), respectively, and $a_{p,i}$ denotes the i-th node or edge of filament p. (2) The paths may be either node-disjoint, edge-disjoint, or unrestricted. (3) The objective of the PCP may be either to obtain a cover of minimum cardinality or minimum weight.

For our purpose, the decomposition of a filamentous network into individual smooth filaments, it seems reasonable to look for an edge-path cover where each edge is covered by (at least) one path and the total (or average) roughness is minimised. Edges that are covered by more than one path naturally correspond to filament overlaps. The minimisation of the average instead of the total roughness favours shorter paths which may be appropriate for some networks.

To define our filament cover problem (FCP) more rigorously, we introduce the roughness r_p of path p and the set $\mathcal{P}$ of all paths in G:

Given a set $\mathcal{E}$ of edges and a set $\mathcal{P}$ of paths with roughnesses r_p, $p \in \mathcal{P}$:

Find a subset $\mathcal{P}_{\text{fil}} \subseteq \mathcal{P}$ with minimal total (or average) roughness R such that each element in $\mathcal{E}$ is covered (at least) once.

The roughness measure r_p of a path p can be chosen arbitrarily and may involve, e.g., the edge weights or the edge alignments. An intuitive choice is the pairwise filament roughness of p (cf. Eq. 1),

$$r_{p,\text{pair}} = \begin{cases} (P-1)^{-1} \sum_{i=1}^{P-1} \left| w_{e_{p,i+1}} - w_{e_{p,i}} \right| & , P > 1 \\ w_{e_{p,1}} & , P = 1 \end{cases}, \tag{S1}$$

where $w_{e_{p,i}}$ denotes the weight of the i-th edge in filament p. The pairwise filament roughness is the average absolute value of the difference between weights of adjacent edges. It reflects the consistency of the edge weights along a filament which is typically smaller within than across filaments (but cf. Discussion). Moreover, if the path consists of a single edge we take its weight as a roughness measure. This choice increases the flexibility of the obtainable filament covers and is necessary to avoid a cover by only individual edges which contribute zero weight when weighted only according the first line in Eq. S1. Another measures for the quality of a filament is the all-to-all filament roughness (cf. Eq. 2)

$$r_{p,\text{all}} = \begin{cases} (P-1)^{-1} \max_{i,j \in \{1,...,P\}} \left| w_{e_{p,i}} - w_{e_{p,j}} \right| & , P > 1 \\ w_{e_{p,1}} & , P = 1 \end{cases}, \tag{S2}$$

which is the average maximal difference between any edge weights in a path p, and again the original weight of the edge is used for a path of length one. Taking into account that most filaments are only moderately bent, we may further wish to minimise the maximal filament deflection angle between adjacent edges of a path p (cf. Eq. 3),

$$r_{p,\text{angle}} = \max_{i \in \{1,...,P-1\}} \tag{S3}$$

$$\text{angle} \left(v_{e_{p,i+1,1}} - v_{e_{p,i+1,0}}, v_{e_{p,i,1}} - v_{e_{p,i,0}} \right)$$

where $v_{e_{p,i,0}}$ and $v_{e_{p,i,1}}$ denote the positions of the start and end nodes of the i-th edge of filament p, respectively. Moreover, $\text{angle}\left(v, v'\right) := \arccos\left(\frac{v \cdot v'}{\sqrt{v \cdot v}\sqrt{v' \cdot v'}}\right)$ is the Euclidean angle of two vectors v and v' and $r_{p,\text{angle}} = 0°$ corresponds to perfectly straight alignment.

SUPPLEMENTAL MATERIAL S2: COMPUTATIONAL INTRACTABILITY OF THE FILAMENT COVER PROBLEM

The FCP is difficult to solve. This is intuitively clear as the number of paths (let alone the number of path covers) increases rapidly with the number of nodes N. Even in planar graphs, the number of closed paths visiting each node once was shown to increase at least exponentially with N [15, 23]. We show now that the FCP is NP-hard, even for planar, cubic graphs. Planar graphs can be drawn on a plane without crossing edges. They are of particular relevance since graphs that are generated from two-dimensional image data are planar by construction [11, 83]. Cubic graphs have only nodes of degree three. A proof of NP-hardness of a problem for planar, cubic graphs directly implies its NP-hardness on general graphs. The basic idea of a typical proof of computational complexity is as follows [40]: A problem of known complexity is selected. By providing a constructive transformation or reduction, a bijection between the known problem and the problem in question is established, i.e., any yes-instance of the decision-version of the known problem is mapped to a yes-instance of the decision-version of the problem of interest and analogously for the no-instances. This reduction proves that the two problems fall into the same class of computational complexity. Our proof is by reduction from the Hamiltonian path problem (HPP) on planar, cubic graphs which is known to be NP-complete [41]. The HPP asks, for a given graph, whether there is a node-path which visits each node exactly once.

First, we note that finding a filament cover on an edge-weighted graph G is equivalent to finding a node-path cover on its node-weighted line graph $L(G)$ (Fig. S1A and B). The line graph $L(G)$ of a graph G has a node of weight w_e for each edge e in G and edges connecting two nodes if the corresponding edges share a node in G.

Second, for a given line graph $L(G)$, we construct a graph such that finding a node-path cover of weight 1 or less is equivalent to solving the HPP. To that end, we add one edge with a terminal node to the line graph and set all original
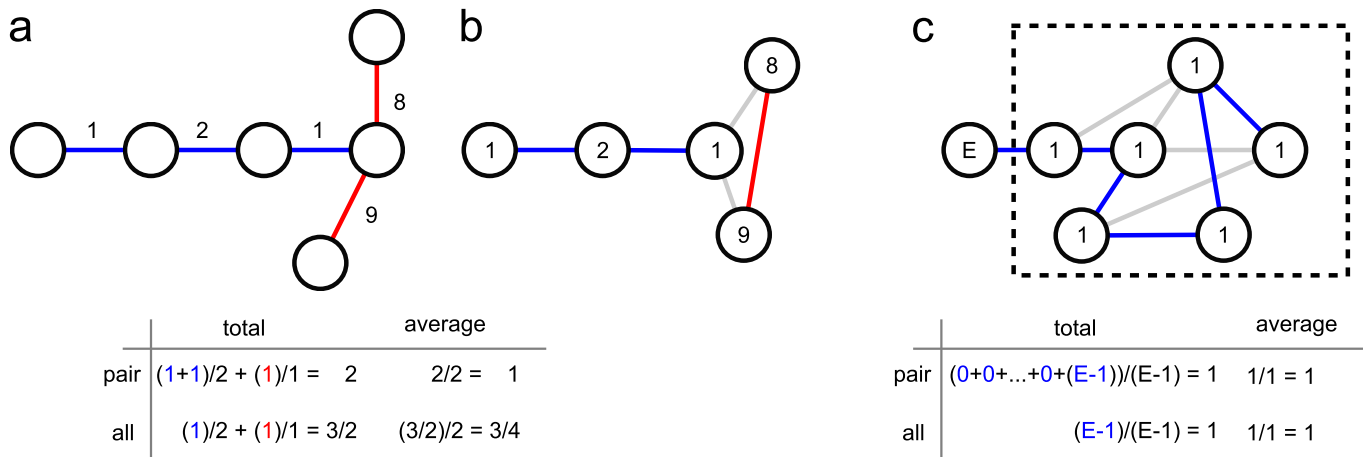
Figure S1. **Proof of NP-hardness of the filament cover problem.** **(a)** Optimal filament cover of an exemplary (edge-weighted) graph. Table with cover roughnesses R for minimisation of total or average roughness and pairwise or all-to-all filament roughness measure, respectively. **(b)** Corresponding (node-weighted) line graph with equivalent path cover and the same roughness results as for the (edge-weighted) graph in (a). **(c)** Extension of an arbitrary graph with node weights 1 by a node of weight E. Here, finding a (node-weighted) path cover of roughness $R = (0 + 0 + \cdots + (E-1))/(E-1) = 1$ or less is equivalent to finding a Hamiltonian path. This equivalence holds for covers minimising the total or average roughness of the cover and using the pairwise or all-to-all filament roughness measure (Eqs. S1 and S2), see table.

node-weights to 1 and the new node-weight to E (Fig. S1C). Then, only a Hamiltonian path ensures a minimal weight of $R = C^{-A} \sum_{i=1}^{C} r_{p_i} = \frac{1}{1} \frac{(0 + \cdots + (E-1))}{(E-1)} = 1$, for both pairwise and all-to-all filament roughness $r_p = \{r_{p,\mathrm{pair}}, r_{p,\mathrm{all}}\}$ (cf. Eqs. S1 and S2) and both minimisation of total and average filament roughness, i.e., $A \in \{0,1\}$.

Finally, we show that finding a Hamiltonian path on a line graph of a planar, cubic graph is NP-complete. It was shown that the HPP is NP-complete on general line graphs via a reduction from the HPP in cubic graphs [14]. This reduction remains valid when planar, cubic graphs are used instead of cubic graphs, for which NP-completeness of HPP is known [41]. Therefore, the decision version of the FCP is NP-complete and the FCP is NP-hard, as claimed. Since the FCP is NP-hard on planar, cubic graphs, it is (at least) NP-hard on general graphs.

SUPPLEMENTAL MATERIAL S3: THE FILAMENT COVER PROBLEM ON TREES IS SOLVABLE IN POLYNOMIAL TIME

While we showed that the FCP is NP-hard on general and even planar, cubic graphs, it is solvable in polynomial time on trees. The polynomial algorithm outlined here is similar to those proposed to find an unrestricted node-path cover where each vertex may be included in multiple paths of minimum cardinality or minimum weight [68].

The basic idea is to assume that a certain path covering a certain edge is in the cover (in a tree, there are at most $N(N-1)/2 = \mathcal{O}(N^2)$ paths to choose from). Upon removal, the tree is split into potentially multiple forests (at most $\mathcal{O}(N)$), each tree of which is decomposed in the same way. The procedure is repeated for each edge (clearly $\mathcal{O}(N)$ in a tree). Thus, this results in a dynamic programming algorithm which has an overall polynomial time complexity of $\mathcal{O}(N^4)$.

The above procedure assumes non-overlapping paths and may be extended to limitedly overlapping paths. For the completely unrestricted case, there would be $\mathcal{O}(2^{\#\mathrm{paths}}) = \mathcal{O}(2^{N^2})$ combinations for covering a given edge to chose from in the first step, and the time complexity of the algorithm would be exponential. However, the problem remains polynomial if we allow only k-fold overlaps, $k = \mathcal{O}(1)$, i.e., each edge may be covered by at most k paths. In the first step of the above algorithm, a given edge may then be covered by at most $\mathcal{O}\left(\binom{N(N-1)/2}{k}\right) = \mathcal{O}(N^{2k})$ edges and consequently the time complexity of the full algorithm is $\mathcal{O}(N^{2k+2})$.

SUPPLEMENTAL MATERIAL S4: APPROXIMATION ALGORITHM FOR THE FILAMENT COVER PROBLEM

Since the FCP is NP-hard even on planar, cubic graphs, we need suitable approximation algorithms. In particular, the approximation algorithms should allow overlapping filaments as well as looped filaments. A natural choice seems to be the formulation of the FCP as a set cover problem (SCP) [56]:

Given an object set $\mathcal{U}$, called universe, and a set $\mathcal{S}$ of sets with costs c_s, $s \in \mathcal{S}$:

Find a subset $\mathcal{S}_{\text{set}} \subseteq \mathcal{S}$ with minimal total (or average) cost such that each element in $\mathcal{U}$ is covered (at least) once.

In our case, the universe corresponds to the set of edges of the given graph ($\mathcal{U} \hat{=} \mathcal{E}$), a set corresponds to a path ($s \hat{=} p$), the cost of a set corresponds to the roughness of a path ($c_s \hat{=} r_p$), and the set cover corresponds to the desired filament cover ($\mathcal{S}_{\text{set}} \hat{=} \mathcal{P}_{\text{fil}}$). We note, that this formulation of the SCP allows overlapping sets, $s \cap s^{'} \neq \emptyset$, $s, s^{'} \in \mathcal{S}$, which directly translates into overlapping filaments in our FCP. By requiring that each element in $\mathcal{U}$ is contained in $\mathcal{S}_{\text{set}}$ exactly once, we may exclude filament overlaps.

An open task is then the generation of a suitable set of paths ($\mathcal{S} \hat{=} \mathcal{P}$). Since for a general graph it is not feasible to find all paths $\mathcal{P}$ (cf. the motivation of the NP-hardness proof of the FCP above), we need to find a representative subset ,$\mathcal{P}^{'}$, of paths. We propose two approaches: (1) We sample paths from $T = 100$ random minimal spanning trees (RMST) of G. To obtain a RMST, each edge is assigned a uniformly distributed random weight and the minimum spanning tree with respect to these weights is computed. Each tree has $N(N-1)/2$ non-trivial, undirected paths that we add to our set $\mathcal{P}^{'}$. However, the paths in a tree cannot contain loops. (2) We perform a modified breadth-first search (BFS) on the nodes, store the generated paths, and stop the search for a path p when it violates a straightness criterion, e.g., $r_{p,\text{angle}} < 60°$ (cf. Eq. S3) which is used throughout the paper. We add all permitted paths to $\mathcal{P}^{'}$. We note that for all real-world filamentous graphs, due to filament thickness, there are spatial constraints on the number of nodes of a graph as well as on the node degrees. Moreover, for the filamentous networks considered here, the radius of curvature of a filament is typically not much smaller than the region of interest. The number of loops is further reduced by the straightness criterion which eliminates paths with a small radius of curvature. Hence, the number of loops in the network is restricted and our heuristically modified BFS allows for loops and yields a representative set $\mathcal{P}^{'}$ in reasonable time.

The SCP may be expressed as a binary fractional linear program [113], and we analogously write the FCP as

$$\text{minimize } \frac{\sum_{p \in \mathcal{P}} r_{p,\text{pair}} x_p}{\left(\sum_{p \in \mathcal{P}} x_p\right)^A} \tag{S4}$$

$$\text{subject to } \sum_{p:e \in p} x_p \geq 1 \text{ for all } e \in \mathcal{E}$$

$$x_p \in \{0, 1\} \text{ for all } p \in \mathcal{P}^{'},$$

where in the first line $A \in \{0, 1\}$ determines whether the total or the average roughness is minimised. In the second line, equality holds for an exact cover. For $A = 0$, Eq. S4 is a binary linear program that may be solved using well-established and efficient algorithms [70, 94].

For $A = 1$, the fractional problem may be rewritten as a binary linear program as well [119, 122]. To that end, we introduce new variables $y = \left(\sum_{p \in \mathcal{P}} x_p\right)^{-1}$ and $z_p = x_p y$, $p \in \mathcal{P}^{'}$. The latter expression is non-linear but may be

replaced by a set of binary linear equations, yielding

$$\text{minimize} \sum_{p \in \mathcal{P}} r_p z_p \tag{S5}$$

$$\text{subject to} \sum_{p: e \in p} z_p \geq y \text{ for all } e \in \mathcal{E}$$

$$\sum_{p \in \mathcal{P}} z_p = 1$$

$$y \geq 0$$

$$y - z_p \leq M - M x_p$$

$$z_p \leq y$$

$$z_p \leq M x_p$$

$$z_p \geq 0$$

$$x_p \in \{0, 1\} \text{ for all } p \in \mathcal{P}.$$

Here, M is a sufficiently large constant that needs to exceed any y (cf. the Big M method [45]). Since $y = \left(\sum_{p \in \mathcal{P}'} x_p \right)^{-1} \leq 1$ for the cover of any non-empty graph, we choose $M = 2$.

Thus, there are a number of options in our FCP: The input set of paths may be obtained by using a modified BFS or from sampling RMSTs or (denoted by either *BFS* or *RMST*). The filaments may overlap or not (*over*/*exact*). The objective of the FCP may be the minimisation of the total or the average roughness (*total*/*avg*). The roughness of a filament may be measured by the pairwise or the all-to-all filament roughness (*pair*/*all*). Solutions of the FCP with different options are compared in the Results.

An implementation of the presented approximation schemes to the FCP with the described options is supplied as an open-source tool, "DeFiNe" (**De**composing **Fi**lamentous **Ne**tworks), under GLP3 at http://mathbiol.mpimp-golm.mpg.de/DeFiNe/. DeFiNe is programmed in Python [112] and employs the packages SciPy [84], NetworkX [46], and cvxopt [30] and PyGTK [36] for a simple and user-friendly graphical user interface. DeFiNe takes as input a weighted graph in the standard .gml file format [50] and outputs a standard .gml graph with filament identities stored as edge colours. Node coordinates may be included in the input file to enable the modified BFS that takes into account edge alignments. Furthermore, manual filament assignments may be included in the input file and the similarity with the automatically obtained filament cover is assessed as described below. In addition, DeFiNe generates a standard, human-readable .csv-table of various individual filament measures for custom analyses. The filamentous structure as well as the manual filament assignments shown in Fig. 1 are available as a .gml file under the above internet address for demonstration purposes.

SUPPLEMENTAL MATERIAL S5: EXTRACTION OF WEIGHTED NETWORKS FROM IMAGES

The procedure used to extract weighted networks from image data is similar to those proposed in [11, 83]: (1) The original grey-scale image are pre-processed to enhance the filamentous structures. Here, a vesselness filter with kernel width of 2 pixels was used for simplicity [37]. (2) In the filtered image, the filamentous structures are separated from the background by applying an adaptive median threshold with a block size of 49 pixels, whereby moderate variations of this size leave our findings largely unchanged. (3) The resultant binary image is skeletonised to obtain the filament centre lines [47]. (4) Then, the nodes of the network under construction are extracted as terminal points, branching points, or crossings of skeleton branches. (5) An edge is inserted between two nodes if they are directly connected via the skeleton. (6) Finally, the edges are weighted by integrating the intensity of the underlying original grey-scale image smoothed with a Gaussian filter with a standard deviation of 5 pixels along the filament and taking its average per unit length of the filament. For the images of the simulated galaxy clusters, the structures obtained by a model-based filter from [104], Figure 6 left middle and bottom rows, are directly employed as binary images and the networks representations are obtained as described above.

SUPPLEMENTAL MATERIAL S6: QUALITY ASSESSMENT OF FILAMENT COVERS VIA STRUCTURE-AWARE PARTITION SIMILARITY MEASURES

The extraction of the filamentous networks from image data enables comparison of the automated filament cover with manual filament assignments. Both, automated cover and manual assignment may be regarded as partitions
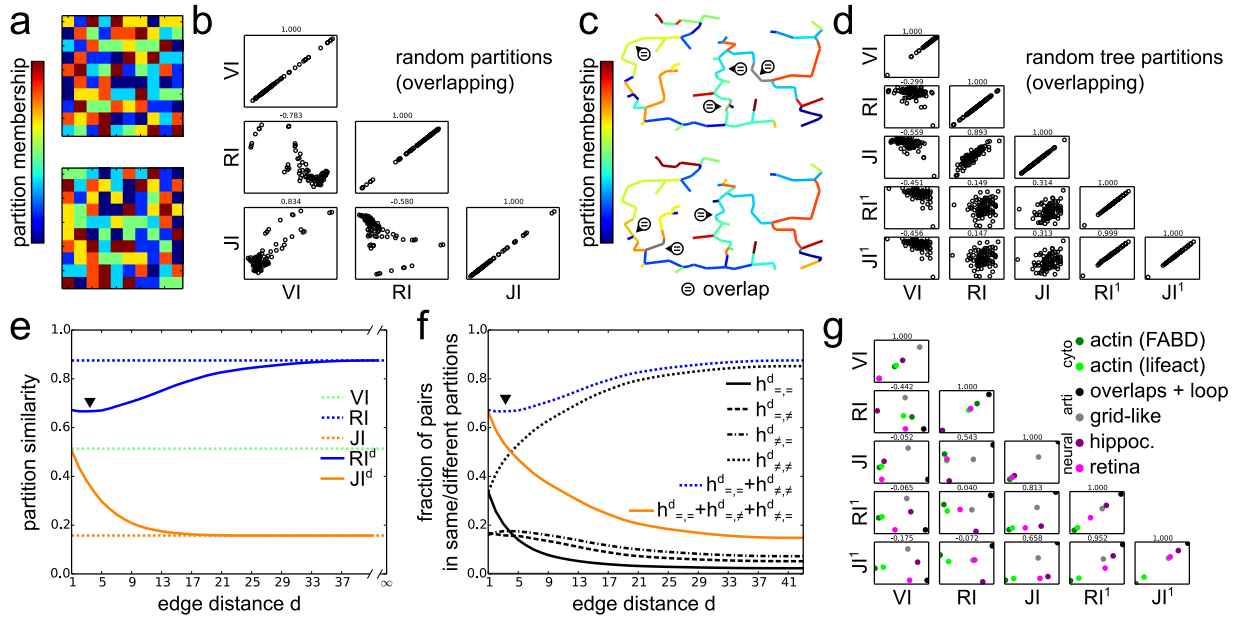
Figure S2. **Comparison of classical and extended partition similarity measures.** Analysis of 100×2 random partitions of sets of 100 numbers plus 10 duplicate ones into $5 - 10$ partitions (a-b). Analysis of a 100×2 path covers of Euclidean minimum spanning trees with 100 nodes distributed uniformly in the unit square, where the paths are drawn randomly and added if the overall overlap of paths is below 10 edges (c-f). Analysis of the similarities between the manual and automated decompositions of the networks studied in the paper (g). **(a)** Colour-representation of two exemplary random partitions as explained above. **(b)** The classical partition similarity measures VI, RI, and JI are not correlated (cf. Kendall rank correlation coefficients $\tau < 0.9$) and may lead to opposing conclusions for the similarity of different partitions. **(c)** Two exemplary random tree path covers with overlaps ($\ominus$). **(d)** The classical partition similarity measures VI, RI, and JI show no correlation among themselves (except for the pairing of RI and JI), nor with the structure-aware RI^1 and JI^1. In contrast, RI^1 and JI^1 are very strongly correlated (cf. $\tau > 0.9$) and yield consistent results for the similarity of different partitions. **(e)** Similarity of the partitions shown in (c) in dependence on maximal distance d between considered pairs of edges (cf. Fig. 3 for a detailed discussion). The RI^d shows a non-monotonous dependency on d (triangle). **(f)** This non-monotonicity of RI^d may be explained by the entries of the contingency table $h_{\times, \times'}$, $\times, \times' \in \{=, \neq\}$. For small distances d, the fraction of true positives (solid black, $h_{=,=}$) drops slower than the fraction of true negatives (dotted black, $h_{\neq, \neq}$) and for larger d, this trend is reversed. Hence, their sum (dashed blue) shows a minimum at intermediate distances d (triangle). In contrast, when summed up (solid yellow), the fast drop in the fraction of true positives dominates over the slightly non-monotony of the false positives and negatives. **(g)** For the investigated artificial and biological networks, the classical measures VI, RI, and JI yield partially opposing results on the similarity of the manual assignment and the automated decomposition (cf. $\tau < 0$). The structure-aware similarity measures RI^1 and JI^1 are strongly correlated and yield consistent results (cf. $\tau = 0.952$).

(where we allow overlapping subsets as well). As measures for the similarity of the automated and manual partitions we use the variation of information, VI, the Jaccard index, JI, and the Rand index, RI, which are commonly used and were shown to estimate similarity reliably for distant and close partitions alike [31, 79, 93]. For given partitions $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_C\}$ and $\mathcal{C}' = \left\{\mathcal{C}'_1, \ldots, \mathcal{C}'_{C'}\right\}$, they are computed via

$$\mathrm{VI}\left(\mathcal{C}, \mathcal{C}'\right) = 1 + (U \log U)^{-1} \cdot \tag{S6}$$

$$\cdot \sum_{i,j} g_{i,j} \left(\log\left(\frac{g_{i,j}}{g_{\cdot,j}}\right) + \log\left(\frac{g_{i,j}}{g_{i,\cdot}}\right)\right),$$

$$\mathrm{RI}\left(\mathcal{C}, \mathcal{C}'\right) = \frac{h_{=,=} + h_{\neq,\neq}}{h_{=,=} + h_{=,\neq} + h_{\neq,=} + h_{\neq,\neq}}, \tag{S7}$$

$$\mathrm{JI}\left(\mathcal{C}, \mathcal{C}'\right) = \frac{h_{=,=}}{h_{=,=} + h_{=,\neq} + h_{\neq,=}}, \tag{S8}$$

where $U = \sum_{i=1}^{C} |\mathcal{C}_i| = \sum_{j=1}^{C'} \left|\mathcal{C}'_j\right|$, $g_{i,j} = \left|\mathcal{C}_i \cap \mathcal{C}'_j\right|$, $g_{\cdot,j} = \sum_{i=1}^{C} g_{i,j}$, and $g_{i,\cdot} = \sum_{j=1}^{C'} g_{i,j}$. The contingency tables $h_{\times,\times'}$, $\times, \times' \in \{=,\neq\}$, provide the numbers of edge pairs which are in the same or different sets in the two partitions, respectively, and is related to $g_{i,j}$ as shown in [52]. All measures are restricted to the unit interval with larger values reflecting higher similarity [78].

While these measures of partition similarity are widely used [63, 79], they pose some difficulties. The variation of information, VI, is only well-defined for disjoint partitions, which occur for non-overlapping filaments. While the Jaccard index, JI, and the Rand index, RI, cover intersecting partitions they may generally yield opposing results. We demonstrate this inconsistency by investigating two types of random partitionings: First, for 100 repetitions, we randomly partitioned 2 sets of 100 numbers and up to 10 duplicates (to simulate overlapping filaments) into $5 - 10$ random partitions (Fig. S2a). While VI and JI were correlated (Fig. S2b; cf. Kendall rank correlation coefficient $\tau > 0$), the other two combinations showed a strong negative correlation (cf. $\tau < 0$). Second, to study filament covers that resemble the decomposition of real filamentous networks more closely, we constructed a relative neighbourhood graph [105, 109] with 100 nodes uniformly distributed in the unit square and computed a random minimum spanning tree (Fig. S2c). For 100 repetitions of this procedure, we partitioned the resultant tree into filaments by choosing a path at random and adding it to the decomposition if the total overlap of any two paths already in the decomposition is below 10 edges (cf. $\ominus$ for overlaps). Again, the correlation among the classical similarity measures was poor or negative (Fig. S2d; except for the correlation between RI and JI; $|\tau| < 0.6$). Although other measure for the similarity of intersecting partitions have been proposed [44, 62, 64], we adhere to RI and JI for simplicity.

More severely, however, the above similarity measures do not take into account the structure of the graph G underlying the (edge-)partitions induced by the obtained filament covers. To date, we are only aware of structure-aware similarity measures for the comparison of partitions whose items are distributed in Euclidean space [7, 26, 123]. Yet, these approaches do not take into account the explicit graph structure of the partitions. To remedy this shortcoming, we introduce a suite of measures, the structure-aware Rand and Jaccard index, RI^d and JI^d, respectively. To that end, the contingency tables $h_{\times,\times'}$ in Eqs. S7 and S8 are replace by distance dependent $h_{\times,\times'}^d$,

$$\mathrm{RI}^d\left(\mathcal{C},\mathcal{C}'\right) = \frac{h_{=,=}^d + h_{\neq,\neq}^d}{h_{=,=}^d + h_{=,\neq}^d + h_{\neq,=}^d + h_{\neq,\neq}^d}, \tag{S9}$$

$$\mathrm{JI}^d\left(\mathcal{C},\mathcal{C}'\right) = \frac{h_{=,=}^d}{h_{=,=}^d + h_{=,\neq}^d + h_{\neq,=}^d} \tag{S10}$$

where $h_{\times,\times'}^d$, $\times, \times' \in \{=,\neq\}$, $d \in \mathbb{N}_{>0}$, count the number of edge pairs which are in the same or different sets in the two partitions, respectively, and which are separated by at most d nodes in G. More precisely, we define

$$h_{\times,\times'}^d = \left\{ \#\left(e_0,e_1\right) \mid e_0 \in \mathcal{C}_i \cap \mathcal{C}'_{i'}, \, e_1 \in \mathcal{C}_j \cap \mathcal{C}'_{j'}, \right. \tag{S11}$$

$$\left. \text{with } i \times j \text{ and } i' \times' j' \text{ and } D_{L(G)}\left(e_0,e_1\right) \le d \right\},$$

where $\#\left(e_0,e_1\right)$ is the number of edges (e_0,e_1) and $D_{L(G)}\left(e_0,e_1\right)$ is the length of the shortest path between nodes in the line graph $L(G)$ of G corresponding to the edges e_0 and e_1. For example, $h_{=,=}^0$ counts the the number of adjacent edges which are in the same set in both partitions (local perspective). In contrast, $h_{\times,\times'}^\infty \equiv h_{\times,\times'}$ reproduce the original measures which do not take into account the positions of edges in the graph (global perspective).

To investigate the performance of our extended, structure-aware partition similarity measures, RI^d and JI^d, we apply them to the artificial graph-based random partitions described above (cf. Fig. S2c). Indeed, when considering the partition membership of neighbouring edges only, i.e., RI^1 and JI^1, the similarity measures yield very consistent results (Fig. S2d; cf. $\tau = 0.999$) in contrast to the lower correlation with the classical similarity measures (cf. $\tau < 0.9$). Investigating the dependency of RI^d and JI^d on the distance d for the tree filament covers shown in Fig. S2d, we find that RI and JI (Fig. S2e; dotted blue and yellow) over- and underestimate the partition similarity with respect to RI^1 and JI^1 (Fig. S2e; solid blue and yellow). Furthermore, we find that RI^d is non-monotonic in d (Fig. S2e; cf. the black triangle). These errors in estimation are explained by the large fraction of false negatives ($h_{\neq,\neq}$) and the small fraction of true positives ($h_{=,=}$), respectively, which dominate for large distances d, i.e., the limit in which the graph structure is ignored (Fig. S2f; dotted black and solid black). Due to the differential increase/decrease of $h_{\neq,\neq}/h_{=,=}$, their combination and therefore RI^1 is non-monotonic (Fig. S2f; dashed blue). Finally, we observe opposing results of the classical partition similarity measures also for the filament covers of artificial and biological filamentous networks investigated in the main text, while our extended, structure-aware measures RI^1 and JI^1 provide consistent similarity results (Fig. S2g).
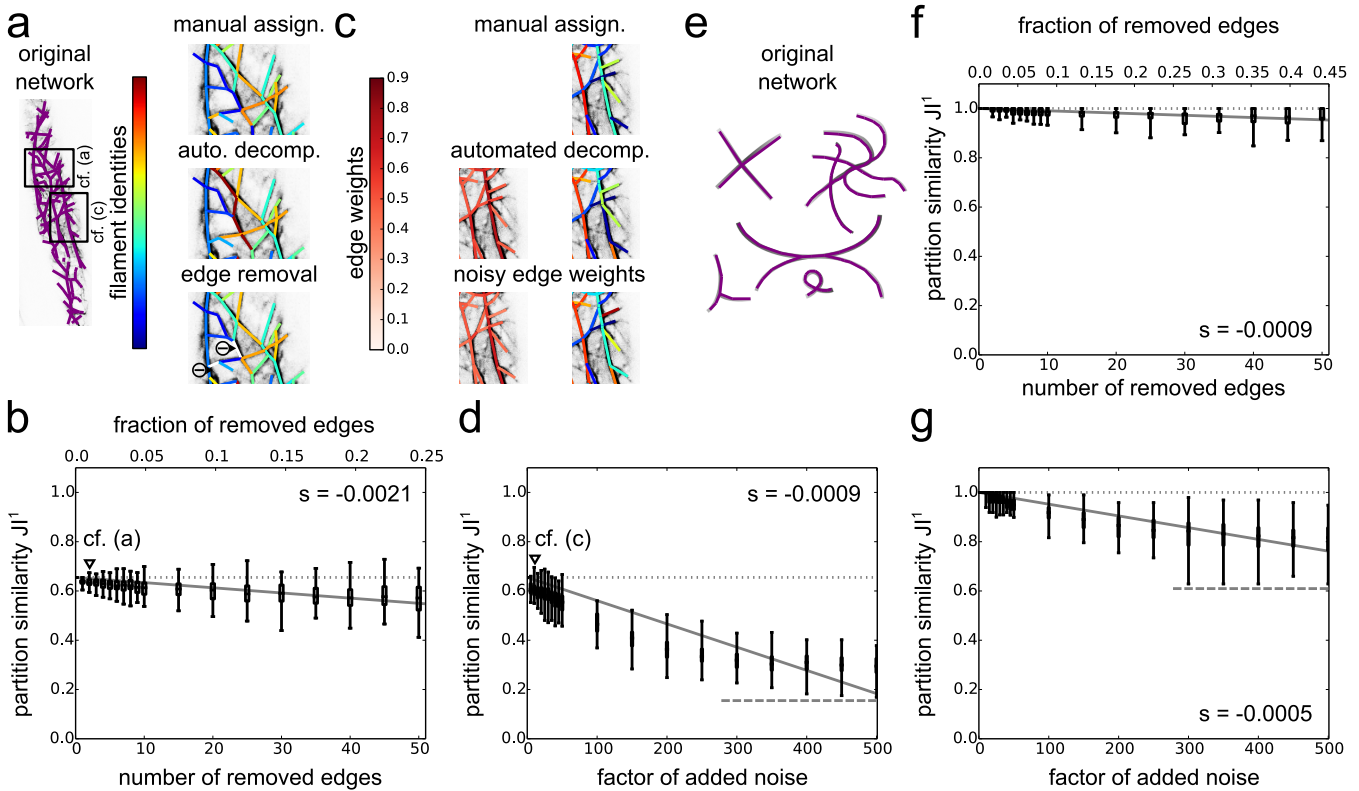
Figure S3. **Analyses of robustness of filament covers against incomplete knowledge of network and image noise.** A cytoskeletal and a contrived network are decomposed automatically by solving the FCP with options given in Fig. 3c and Fig. 1c, respectively. **(a)** Overlay of extracted actin network structure and original image data (left panel). Sections of cytoskeletal network with edge colours representing the manual assignment, the optimal filament cover obtained for the full, non-disrupted network, and the optimal filament cover after removal of two edges which are shown in white (right panels). **(b)** Similarity of manual filament assignment and automated filament covers after removal of increasing numbers of edges, measured by structure-aware Jaccard index JI^1. On average, JI^1 decreases with the number of removed edges as shown by a linear fit with slope $s = -0.0021$ (solid grey line). Occasionally, the removal of edges increases the accuracy of the filament cover above the accuracy of the original solution (dotted grey line and triangle; cf. panel (a)). **(c)** Sections of cytoskeletal network with edge colours representing the original edge weights and the edge weights after adding Gaussian noise (left panels). Sections of cytoskeletal network with edge colours representing the manual assignment, the optimal filament cover obtained for the full, non-disrupted network, and the optimal filament cover after adding Gaussian noise (right panels). **(d)** Similarity JI^1 of manual filament assignment and automated filament covers after adding Gaussian noise. On average, JI^1 decreases with increasing noise factor as shown by a linear fit with slope $s = -0.0009$ (solid grey line). Occasionally, the noisy edge weights lead to an increase in accuracy of the filament cover above the accuracy of the original solution (dotted grey line and triangle; cf. panel (c)). The decrease levels off for large noise factors and JI^1 approaches a constant value (dashed grey line). **(e)** Overlay of extracted contrived network structure and original image data. **(f)** Results for the contrived network analogue to those presented for the cytoskeletal network in panel (b). The average change in JI^1 per removed edge is captured by a linear fit with slope $s = -0.0009$. **(g)** Results for the contrived network analogue to those presented for the cytoskeletal network in panel (f). The average change in JI^1 per unit increase in the noise factor is captured by a linear fit with slope $s = -0.0005$.

SUPPLEMENTAL MATERIAL S7: ROBUSTNESS OF FILAMENT COVERS AGAINST INCOMPLETE KNOWLEDGE OF UNDERLYING NETWORK STRUCTURE AND IMAGE NOISE

Our approach enables accurate decomposition of a given filamentous network into its constitutive filaments (cf. Results). However, the preceding extraction of the network from image data is often non-trivial (cf. Methods). Therefore, to assess the robustness of our approach, we test how the accuracy of our filament decomposition is affected (1) by incomplete knowledge of the true underlying network structure and (2) by image noise which affects the edge weights of the extracted network. We perform these analyses for the actin cytoskeleton shown in Fig. 3 (Fig. S3a, left panel) and the contrived network shown in Fig. 1 (Fig. S3e).

(1) First, we start from the original, weighted network and randomly remove one of the E edges to model erroneous

segmentation. For the disrupted network, we recompute the optimal filament cover (with the same options as in Figs. 1c and 3c, respectively) and calculate its agreement with the original manual segmentation (measured by the structure-aware Jaccard index JI^1; the removed edge is assigned a dummy label). We repeat the procedure for E networks from which a single, randomly chosen edge has been removed. Next, we repeat the procedure for E networks from which a randomly chosen double of edges has been removed. We then proceed with triplets, quartets, and so on up to subsets of 50 randomly chosen edges.

As expected, the removal of increasing numbers of edges typically decreases the agreement of the automated filament cover with the manual assignment for the cytoskeletal as well as the contrived network (Fig. S3b and f). For both networks, however, the decrease is slow and JI^1 increases only by around 0.002 per removed edge (cf. Fig. S3b and f, solid grey line indicates linear fit). Interestingly, for the actin cytoskeleton, the removal of certain edges may even increase the accuracy of the filament cover (Fig. S3a, right panels show manual filament assignment and automated filament cover the original network, and an exemplary filament cover obtained after the removal of two edges, coloured white here, which improves the agreement with the manual assignment; cf. Fig. S3b, dotted grey line and triangle).

(2) Second, we simulate image noise by adding centred Gaussian noise Δw to the edge weights of the original network with

$$\mathrm{E}\left[\Delta w\right] = 0, \tag{S12}$$

$$\mathrm{Sd}\left[\Delta \mathrm{w}\right] = \left(1 + \frac{f}{100}\right) w. \tag{S13}$$

We normalise the standard deviation of the added noise by the original edge weights to avoid extreme fluctuations, and f is referred to as noise factor. For each noise factor, we construct 100 networks, recompute the optimal filament covers, and measure their agreement with the manual filament assignment, as in the first scenario above.

For both the contrived and the cytoskeletal network, the accuracy of the filament cover decreases with increasing noise, as expected (Fig. S3d and g). However, this decrease in accuracy is slow and JI^1 decreases by less than 0.001 when increasing the standard deviation of the noise by 1% of the original edge weights, i.e., when increasing the noise factor by one (cf. Fig. S3d and g, solid grey lines indicate linear fits). We note that with increasing edge noise the accuracy of the filament cover approaches a constant, non-zero JI^1 which reflects that some information about the filament structure maybe obtained from the topology of the network alone, irrespective of the edge weights (cf. Fig. S3d and g, dashed grey lines).

SUPPLEMENTAL MATERIAL S8: FILAMENT ANALYSIS FOR NETWORKS EXTRACTED FROM MOVIE OF PLANT ACTIN CYTOSKELETON

To further strengthen our statistical analyses of cytoskeletal actin filaments (cf. Fig. 3), we investigate a complete movie of a plant cytoskeleton of 100 frames over 200 s (cf. Methods for details). For each frame, we extract a weighted network representation of the cytoskeleton as described above (cf. Methods for details) and solve the FCP with options described in Fig. 3, i.e., we solve the exact FCP (*exact*) for paths from a modified breadth-first search (*BFS*) and by minimising the total (*total*) pairwise filament roughness (*pair*). Analysis of various properties of the automatically obtained filaments confirms our findings in Fig. 3: The filaments show a preferential alignment parallel to the cell axis throughout the movie (Fig. S4a). The distribution of filament lengths, pooled across the duration of the movie, confirms the reported gamma distribution (Fig. S4b; maximal likelihood fits of normal, Weibull, and Rayleigh distributions yield higher values for the Akaike information criterion [1]). Filament length is correlated with filament weight, i.e., longer filaments are typically thicker (Fig. S4c; Pearson correlation p-value $p_\mathrm{P} < 0.05$). Moreover, the correlation between different measures of filament curvedness, i.e., the filament bending and the maximal filament angle, are consistently negatively correlated with the filament length (Fig. S4d; $p_\mathrm{P} < 0.05$).

In addition to these previously analysed features of filamental organisation, we study the course of different filament properties over time: The average filament weight shows large fluctuations and is non-stationary over the recording period (Fig. S4e; cf. augmented Dickey-Fuller test p-value $p_\mathrm{ADF} \geq 0.05$). This non-stationarity suggests substantial changes in the prevalence of fine actin filament and thick bundles, respectively, and prompts further investigations. However, we found that the average filament length as well as the average filament bending remain stationary over the course of 200 s (Fig. S4e; cf. $p_\mathrm{ADF,length} < 0.05$ and $p_\mathrm{ADF,conv} < 0.05$). Since the length distribution of filaments tunes the mechanical properties of filamentous networks [8, 57], this stationarity of the average filament length may be of immediate biological relevance. The stationarity of the average filament bending may be a direct consequence of the roughly constant filament length distribution (cf. Fig. S4d) in combination with the resultant physical constraints of actin filament length on filament bending.
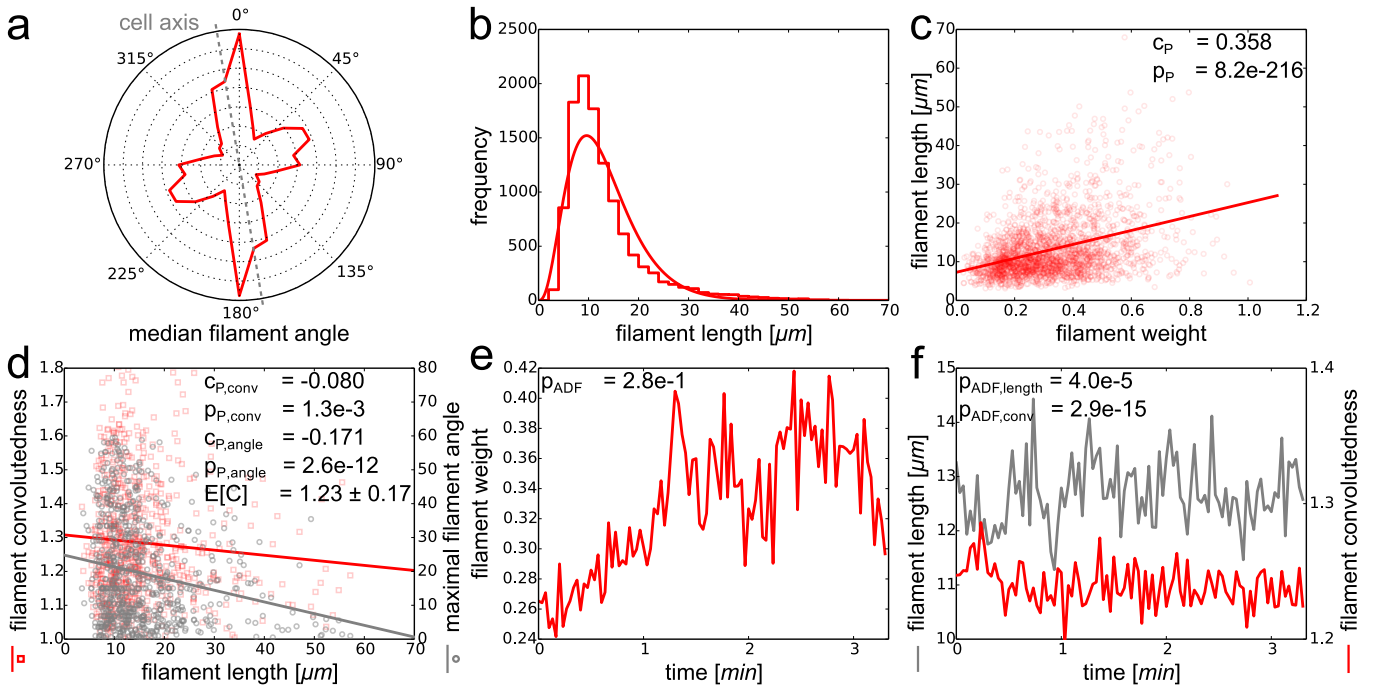
Figure S4. Filament analyses of 100 **cytoskeletal networks.** Results from filament decompositions of 100 cytoskeletal networks extracted from a movie of a plant cytoskeleton over 200 s. The cytoskeletal networks are decomposed automatically by solving the exact FCP (*exact*) for paths from a modified breadth-first search (*BFS*) and by minimising the total (*total*) pairwise filament roughness (*pair*; cf. Fig. 3). **(a)** The distribution of median filament angles shows that the majority of filaments is aligned parallel to the cell axis (grey dashed line). **(b)** Filament lengths (bars) follows a gamma distribution (line shows maximum likelihood fit). **(c)** Filament length correlates with filament weight (cf. linear regression and Pearson correlation coefficient $c_P > 0$ and p-value $p_P < 0.05$) **(d)** Scatter plot of filament convolutedness versus filament length shows a negative correlation (cf. red squares, $c_{P,\mathrm{conv}} < 0$, and $p_{P,\mathrm{conv}} < 0.05$) with an average convolutedness of $\mathrm{E}[C] = 1.23 \pm 0.17$. The maximum filament angle correlates negatively with the filament length (cf. grey circles, $c_{P,\mathrm{angle}} < 0$, and $p_{P,\mathrm{angle}} < 0.05$), indicating that longer (and thicker, cf. (c)) filaments are less curved. **(e)** Time series of average filament weight over 200 s shows large fluctuations and is non-stationary (cf. augmented Dickey-Fuller test p-value $p_{\mathrm{ADF}} \geq 0.05$). **(f)** Time series of filament length and convolutedness are stationary over the recording period (cf. $p_{\mathrm{ADF,length}} < 0.05$ and $p_{\mathrm{ADF,conv}} < 0.05$).

SUPPLEMENTAL MATERIAL S9: OVERVIEW OF DIFFERENT STAGES OF FILAMENT DECOMPOSITION OF ARTIFICIAL, BIOLOGICAL, AND COSMIC NETWORKS

We test our method of decomposing a given weighted network into filaments by solving the FCP for different filamentous networks. In addition to the four networks presented in the main text and the 100 frames analysed in Supplemental Material S8, we investigate four more networks of different types and show the different stages of our analysis. Starting from grey-scale image data of contrived, neural, cytoskeletal and cosmic network structures (Fig. S5, 1st column), we pre-process the images to obtain a binary representation of the filament centre lines (Fig. S5, 2nd column), and extract a weighted network representation as described in the Methods (Fig. S5, 3rd column). For the contrived and biological and the cosmic networks, we manually assign filament identities and compute the connected components, respectively (Fig. S5, 4th column). Finally, we decompose the networks into filaments by solving the FCP with different options (Fig. S5, 5th column). For the first contrived network, we allow overlapping filaments (Fig. S5a) while for the second, grid-like contrived network (Fig. S5b), the neural networks (Fig. S5c and d), the cytoskeletal networks (Fig. S5e and f), and the cosmic webs (Fig. S5g and h), we obtain exact filament covers with options described in Fig. 1e. The agreement of manual assignments and automated filament decompositions of the studied networks is measured by the classical and the structure-aware Jaccard indices JI and JI[1] and shows good agreement (JI[1] close to 1, and cf. discussion of Fig. 3d) despite occasional over- (cf. $\oplus$) or under-segmentation (cf. $\ominus$) of filaments.

SUPPLEMENTAL MATERIAL S10: OPEN CONTOUR-BASED FILAMENT DECOMPOSITION AND FILAMENT COVER-BASED POST-PROCESSING

Finally, we demonstrate how our filament cover-based approach may be used to post-process and improve filament decompositions obtained from other, e.g., open contour-based approaches. For the demonstration, we select SOAX [121], a fully automated, stretching open active contour-based approach which is available as an open-source software tool to extract a network-like representation (i.e., coordinates of filament centre lines as well as junctions are provided) from image data. As a test case, we study the contrived filamentous structure investigated in Fig. 1. For a fair comparison of our and the open contour-based approach, we apply SOAX to the pre-processed and segmented image data (cf. Methods and Fig. S6a, second panel) to which we further apply a Gaussian filter of unit standard deviation to obtain smooth intensity gradients required by the algorithm. SOAX is run using the default parameters and the resulting filament identities are manually assigned to match those of the manual solution (Fig. S6b). To quantify the quality of the decomposition, we manually assign filament identities in our original network representation (cf. Fig. S6a, third panel) according to the open contour-based result (cf. Fig. S6b) and compare the result to the manual assignment (cf. Fig. S6a, fourth panel). The structure-aware Jaccard index $JI^1 = 0.938$ is close to 1 and indicates good agreement between open-contour based decomposition and manual filament assignment. We note that some junctions/nodes obtained from SOAX are split in two in comparison to our extracted networks (cf. intersecting $\bigcirc$).

Moreover severely, some filaments are over-segmented and thus fragmented (cf. $\oplus$), especially overlapping filaments which are not captured in the open contour-based approach (cf. $\ominus$). To remedy this shortcoming, we apply our filament cover-based approach to post-process the open contour-based decomposition and merge over-segmented filament fragments. To this end, we convert the open contour-based filament representation into a weighted network, where edge weights represent average filament segment intensities as before (cf. Methods and Fig. S6c). As before, a collection of paths $\mathcal{P}'$ is sampled using a breadth-first search (BFS) and their pairwise roughness values r_p, $p \in \mathcal{P}'$, are computed according to Eq. S1 ($pair$). Then, to take into account the initial open contour-based filament decomposition $\mathcal{F}$ as a starting point, in which certain edges have already been assigned to the a given filament, we modify the roughness values of the sampled paths: For each initial filament or fragment that is fully contained within a sampled path, the roughness of that path is decreased by a large value, $R_{\text{filament}} = 10^4$, which is larger than any r_p to favour the inclusion of these filaments or fragments in the optimal filament cover. Since the subtraction of R_{filament} yields negative roughness values which would lead to the inclusion of all these paths, we add another, even larger constant $R_{\text{offset}} = 10^8 > R_{\text{filament}}$ to all roughness values, i.e.,

$$r'_p = r_p - \sum_{\substack{f \in \mathcal{F} \\ f \subset p}} R_{\text{filament}} + R_{\text{offset}}. \tag{S14}$$

For these modified roughness values r'_p, we solve the FCP by minimising the total roughness ($total$) and allowing for overlaps ($over$; Fig. S6d). The resulting post-processed filament decomposition merges several filament fragments which were over-segmented by the open-contour based approach and shows very good agreement of $JI = 0.776$ and $JI^1 = 1.000$ with the manual filament assignment. Interestingly, in this decomposition, parts of two filaments are interchanged (cf. $\oplus$) as in Fig. 1f for different FCP options. In conclusion, for any approach that detects filaments from image data and yields a weighted network representation, our filament cover-based approach may provide a helpful means to further post-process and enhance the accuracy of the obtained filament decomposition.

edge weight

filament identities

a original image   segmented image   weighted network   manual assignment   BFS/over/pair/total
JI = 1.000
JI¹ = 1.000

b   BFS/exact/pair/total
JI = 0.742
JI¹ = 0.872

c   JI = 0.427
JI¹ = 0.937

d   JI = 0.397
JI¹ = 0.883

e   JI = 0.366
JI¹ = 0.655

f   JI = 0.193
JI¹ = 0.701
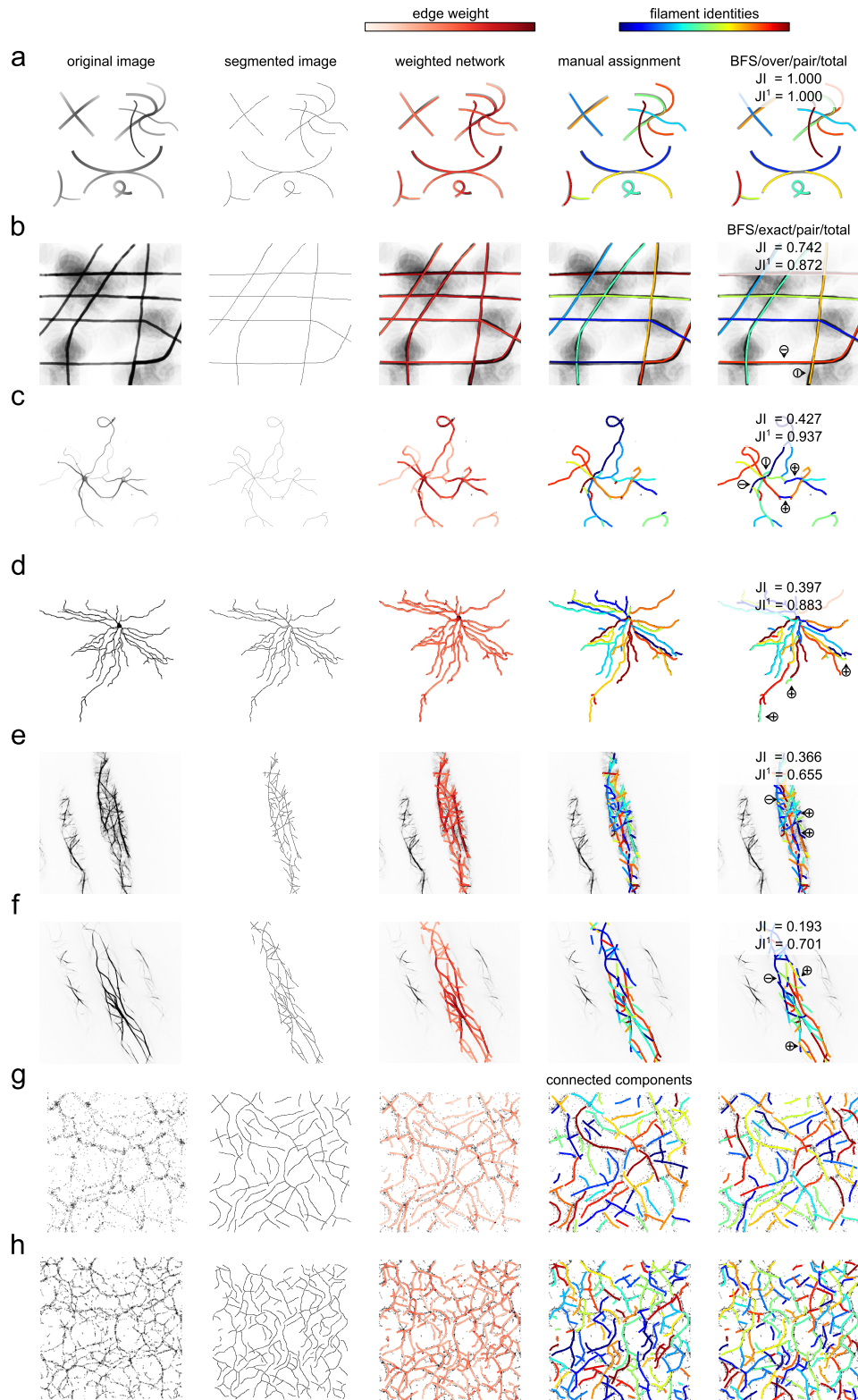
g   connected components

h

Figure S5. **Overview of studied networks, manual assignments, and filament covers obtained from solving the FCP.** Original grey-scale image data (1st column), binary images of filament centre lines (2nd column), extracted networks with colour-coded edge weights (3rd column), manual filament assignments of contrived and biological networks and connected connected components of cosmic networks, respectively (4th column), and automatically obtained filament covers (5th column). Agreement between manual decompositions and automated filament cover is quantified by a number of measures (cf., e.g., Methods and Fig. 2), here the classical and the structure-aware Jaccard indices JI and JI^1 are shown. **(a)** Contrived network with crossing and overlapping filaments and a loop (cf. Fig. 1). **(b)** For a contrived, grid-like network, the automated decomposition correctly detects most of the filaments (JI^1 close to 1). Only the filament in the bottom right corner with a kink is over-segmented ($\oplus$) because the curvature restriction of the initial paths does not allow such large angles (about $90°$ here, cf. Eq. S3). **(c)** Neural network of hippocamal neuron (cf. Fig. 2). **(d)** The decomposition of the network of a retinal ganglion cell shows good agreement with the manual results (JI^1 close to 1). A few filaments are over-segmented ($\oplus$), e.g., due to kinks in the filaments that are not captured by the initial set of paths (cf. the centre $\oplus$). **(e)** Cytoskeletal network of actin filaments (cf. Fig. 3). **(f)** For the actin network extracted from the confocal recording of a Lifeact-labelled cytoskeleton, the automated partitions agrees well with the manual results (JI^1 close to 1, and cf. discussion of Fig. 3d). A few examples of over- and under-segmentation ($\ominus$) are marked. **(g)** Cosmic web of galaxies (cf. Fig. 4). **(h)** The dense web of simulated galaxies consists of many connected components that are further decomposed into filaments (cf. Fig. 4 for a discussion). Image data for panels (g) and (h) from: Stoica et al., A&A, 434, 423-432, 2005, reproduced with permission © ESO.
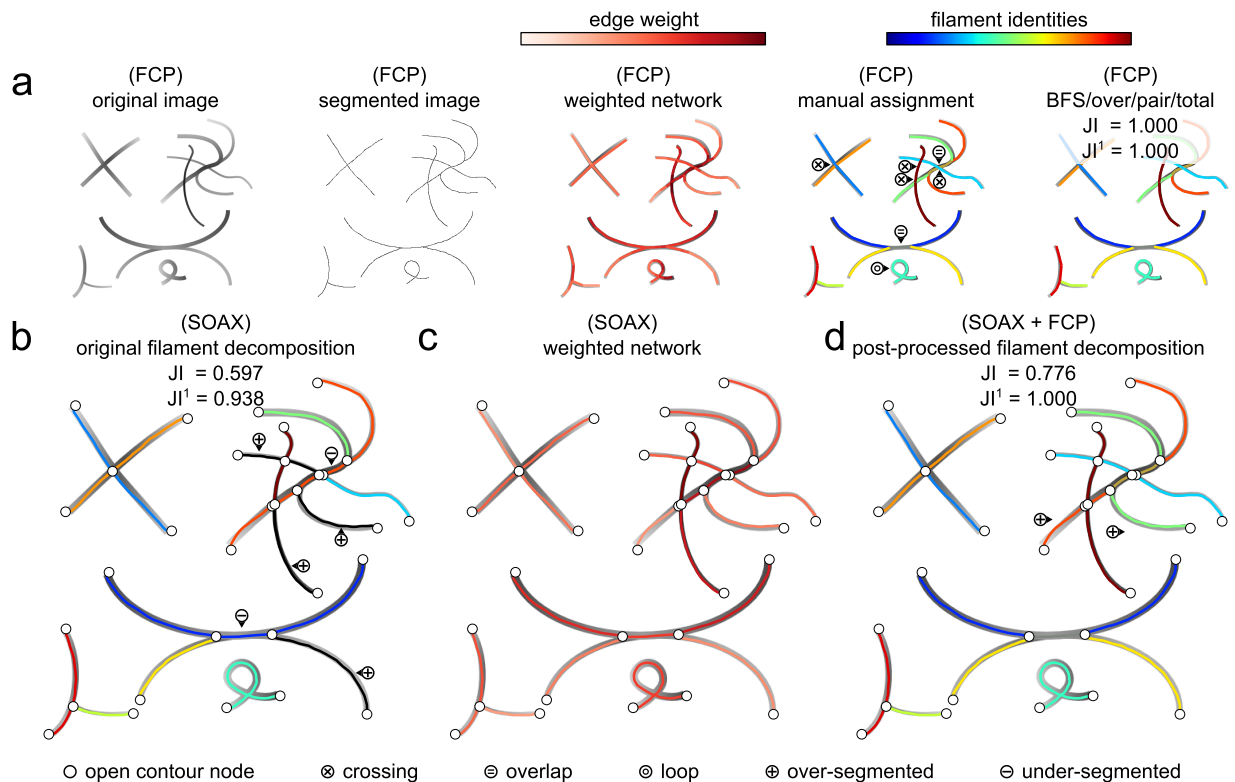

Figure S6. **Open contour-based filament detection and filament cover-based post-processing. (a)** Different stages of our filament cover problem (FCP)-based analysis for a contrived filament structure, from original image to segmented filament centre lines and weighted network representation, manual filament assignment and automated solution (cf. Fig. S5a for further explanations). **(b)** Filaments and junctions (cf. circles) identified from the segmented filament centre line image using SOAX, an stretching open active contour-based approach [121]. Colour-coded filament identities were manually assigned to match those of the manual solution in (a) and excess filament fragments were coloured black. While the agreement with the manual solution is good (JI^1 close to 1), some filaments are over-segmented (cf. $\oplus$) and thus fragmented, especially at locations of filament overlaps (cf. $\ominus$). **(c)** Weighted network representation of the contrived filamentous structure obtained from SOAX. **(d)** Using the filament assignments from SOAX in (b) as a starting point, our filament cover-based approach is used to post-process the filament decomposition, which merges broken filament fragments and improves the agreement with the manual solution ($JI^1 = 1$).