Advances in Geo-Energy Research

Vol. 20, No. 2, p. 101-113, 2026

Original article

Topological data analysis for pore-network extraction in porous media

Jie Liu[1,2], Tao Zhang[3], Zhipeng Duan[4], Yuze Zhang[4]*, Shuyu Sun[1]*

[1]*School of Mathematical Sciences, Tongji University, Shanghai 200092, P. R. China*

[2]*Physical Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia*

[3]*Institute of New Energy, China University of Petroleum (East China), Qingdao 266580, P. R. China*

[4]*School of mathematical Sciences, Ministry of Education key Laboratory of NSLSCS, Nanjing Normal University, Nanjing 210023, P. R. China*

Keywords:
Pore-network model
topology data analysis
string method
porous media

Cited as:
Liu, J., Zhang, T., Duan, Z., Zhang, Y., Sun, S. Topological data analysis for pore-network extraction in porous media. Advances in Geo-Energy Research, 2026, 20(2): 101-113.
https://doi.org/10.46690/ager.2026.05.01

Abstract:
Pore-network models are widely used to describe pore-scale flow in porous media, and their reliability depends critically on accurate extraction of pore and throat structures. A new extraction framework, termed the topological pore-network finder, is proposed in this work, which combines topological data analysis, medial access path search, and flashlight search medial axis. The topological data analysis is used to identify pore connectivity and cluster the void space, thereby providing robust initial pore centers. The medial access path search method then traces strings between connected pore centers along the medial axis, while the flashlight search medial axis method is used to refine the resulting paths and improve computational efficiency. The method is validated using toy porous media, two- and three-dimensional digital rock samples. Sensitivity analyses show that the pore-network finder is stable with respect to image resolution and string discretization. Compared with the classical maximal-ball method, the pore-network finder achieves at least an order-of-magnitude acceleration while preserving the main geometric statistics and flow-response characteristics of the extracted networks. In addition, because the method operates in continuous space and can reuse information from previous states, it is well suited to quasi-dynamic updates during deformation. The pore-network finder therefore provides an efficient and accurate tool for pore-network extraction and subsequent pore-scale characterization in geo-energy systems.

1. Introduction

Understanding pore-throat architecture in subsurface rocks is fundamental to geo-energy systems because fluid storage, transport, and recovery in geological formations are governed by pore connectivity, permeability, and structural heterogeneity (Cai et al., 2017; Feng et al., 2024). Accurate characterization of pore-throat networks is therefore essential for predicting fluid flow and transport in porous media and for a wide range of geo-energy applications (Shao et al., 2023; Chen et al., 2025), including petroleum engineering and hydrology (Liu et al., 2022).

Pore-network modeling is one of the most effective approaches for simulating fluid transport in porous media once a representative network structure has been extracted (Zhang et al., 2015; Qin et al., 2024). The detailed pore geometry is reduced to a network of nodes and edges, where nodes represent pores and edges represent throats or pore connections (Cui et al., 2022). By simplifying the underlying geometry (Thompson and Fogler, 1997; Chareyre et al., 2012; Morimoto

Yandy Scientific Press

*Corresponding author.
E-mail address: jie.liu.1@kaust.edu.sa (J. Liu); tao.zhang@upc.edu.cn (T. Zhang); zhipeng@njnu.edu.cn (Z. Duan); 05439@njnu.edu.cn (Y. Zhang); suns@tongji.edu.cn (S. Sun).

Received January 26, 2026; revised February 28, 2026; accepted March 16, 2026; available online March 19, 2026.

et al., 2022), pore-network extraction enables much more efficient analysis of flow and transport than direct simulation in the full pore space (Raoof and Hassanizadeh, 2010).

A variety of pore-network extraction (PNE) methods have been developed, each with distinct computational characteristics (Jiang et al., 2017; Ni et al., 2017). The medial axis method identifies the skeleton of the pore space by locating points that are equidistant from the solid boundaries (Lindquist et al., 1996). This approach involves thinning or burning the pore space, which reduces the pore structure to a low-dimension representation without changing the topology of the pore network. This simplification allows the pore network to be directly modeled. However, it can sometimes produce complex and noisy skeletons, especially in irregularly shaped pores (Lee et al., 1994). The classical maximal-ball method instead fits the largest admissible spheres inside the void space (Silin and Patzek, 2006). The centers of these spheres represent the pore centers, between them the minimal spheres define the throats.

Both medial axis and maximal ball approaches are fundamentally pixel- or voxel-based and therefore depend strongly on image resolution. Low image resolution can introduce noise into pore-structure extraction, obscure fine features, and compromise the fidelity of the reconstructed network, whereas high image resolution imposes greater demands on computational resources. To alleviate the inherent trade-off between accuracy and efficiency in pixel-based pore-network extraction, pixel-free methods have been developed. The flashlight search medial axis (FSMA) method achieves relatively high extraction accuracy, but its step-by-step tracing of the pore-space medial axis limits computational efficiency (Liu et al., 2024b). The medial access path search (MAPS) method was later introduced as an optimized extension of the FSMA, further improving the efficiency of pixel-free searching by incorporating acceleration strategies and the energy landscape (Zhang et al., 2024). Nevertheless, both methods remain limited in their ability to robustly resolve connectivity relationships among pores.

The topological data analysis (TDA), by contrast, provides a mathematically rigorous way to characterize connectivity and shape. Its main tools include persistent homology and the Mapper algorithm. Persistent homology tracks the emergence and persistence of topological features across scales (Carlsson, 2009; Edelsbrunner and Harer, 2022). The Mapper algorithm (Singh et al., 2007), inspired by Morse theory, constructs a simplicial complex that captures the essential connectivity of a dataset. Because pore-network structures are naturally represented as node-edge graphs, the TDA offers an attractive route to move beyond the limitations of pixel-free extraction methods.

In this study, a new pore-network extraction algorithm is proposed, the topological pore-network finder (TPF). By moving beyond conventional pixel-based representations, TPF provides a faster framework for pore-network extraction and is particularly well suited to tracking pore-network evolution during dynamic compression. Its robustness and applicability are demonstrated through sensitivity analyses, real rock images, and three-dimensional digital rock examples.

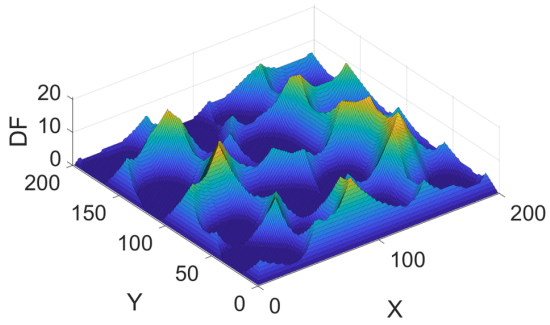

Fig. 1. DF landscape in porous media, where the DF is zero in the solid phase.

2. Methodology

2.1 Flashlight search and medial axis path methods

To facilitate the discussion, a two-dimensional (2D) example is shown in Fig. 1, where the distance function (DF) is defined for all void pixels (Liu et al., 2024b):

$$D(v,S) = \min\{D(v,s),\ s \in S\} \tag{1}$$

where D denotes the minimal distance from a void point to the solid phase; v and s denote points in the void and solid domains, respectively; and S denotes the solid region. In three-dimensional (3D) space, the resulting landscape can be viewed as a mountain surface, where peaks correspond to pore centers and saddles correspond to pore throats. The medial line of the pore space is associated with non-smooth locations of the DF, which provides a criterion for identifying the medial axis (Niasar et al., 2009).

Pixel-free methods identify pore networks directly in continuous space by searching for the medial structure, rather than by inferring connectivity from pixel counting. In the previous work, the FSMA method has been proposed, which performs medial-axis searching in continuous space (Liu et al., 2024b). The FSMA uses gradient descent to locate an initial pore center. Once this center is identified, a search region is constructed, a curve in 2D or a surface in 3D. Within this region, the gradient and Laplacian of the DF are used to detect points on the medial axis, because the DF is typically non-smooth there and its gradient becomes discontinuous. As the search proceeds along the medial axis, pore throats are identified as local minima of the DF encountered before the next pore center, whereas pore centers correspond to local maxima. The search direction is determined by following the maximum gradient (Liu et al., 2024b):

$$\max\left[\frac{D(v,s)-D(v+(Rcos\theta_i,Rsin\theta_i),S)}{R}\right],\ i=1\ldots M \tag{2}$$

where R is the search radius and $\theta_i=(i-1)2\pi/M$, M is the number of discretization points within R. Although the FSMA is effective for locating pore centers, it advances step by step and is therefore not ideal for parallel computation.

The MAPS method is another pixel-free approach for extracting pore-network structures (Zhang et al., 2024). Starting from an initial string connecting two pore centers, the string

method (E et al., 2002) is used to evolve the string toward the medial line of the pore space. This procedure is analogous to identifying a minimum-energy path (Zhang et al., 2023), if the distance landscape is interpreted as an energy landscape. The DF can then be transformed into an energy function F as follows (Zhang et al., 2023):

$$F(v,\ S) = -D(v,\ S) \tag{3}$$

In the energy landscape, the minimum-energy path is the curve γ connecting two pore centers, which are treated as minimum local energy. The tangent of the path can then be written as follows (Zhang et al., 2023):

$$F\nabla F_{sys}(\gamma) - (\nabla F_{sys}(\gamma), \hat{\gamma})\hat{\gamma} = 0 \tag{4}$$

where F_{sys} denotes the system energy and $\hat{\gamma}$ is the unit tangent vector of the path. The same relation can also be written as:

$$(\nabla F_{sys}) \perp (\gamma) = 0 \tag{5}$$

Each string can be calculated independently, which makes the MAPS well suited to parallel computation once the connectivity information has been established during preprocessing.

Both the MAPS and FSMA require prior topological connectivity between pores to determine the appropriate extraction. If this information is available in advance, unnecessary calculations can be avoided, and the MAPS becomes much more efficient. The TDA provides a natural framework for extracting these essential connectivity relations from complex datasets. More broadly, the TDA uses tools from algebraic topology to characterize the shape of data in a manner that is robust to noise.

2.2 Mapper algorithms in TDA

To identify the holes and tubes in a dataset before selecting initial points for the PNE, the global structure of the data must first be characterized. The Mapper algorithm in TDA provides an effective way to achieve this by offering both a compact representation and a visual description of the underlying topology (Singh et al., 2007).

The core idea of the Mapper algorithm is a generalization of the Reeb graph (Biasotti et al., 2008). Given a function $f: M \to \mathbb{R}$, each point in the Reeb graph corresponds to a connected component of the preimage of a value under f. The Reeb graph therefore serves as a topology-preserving dimensionality-reduction tool. Mapper extends this idea by using a filter function to extract the key global topological features of a dataset and represent them as a graph or simplicial complex. A simplicial complex is a union of simplices of different dimensions, such as points, line segments, triangles, and their higher-dimensional analogues. It provides a combinatorial representation of space. Such complexes not only visualize essential connectivity but also encode rich topological information in an algebraic form.

Before delving into the core concepts of the Mapper algorithm, it is important to revisit the notions of covers and nerves in algebraic topology. For a given topological space X, the covering is a collection of open sets $U = \{U_i\}_{i\in I}$ in X such that their union equals X, i.e., any point in X will belong to some open sets in U. Let ε be a positive real number. A usual covering for $\mathbb{R}$ is $U_\varepsilon = \{U_i\}_{i\in\mathbb{Z}}$ where $U_i = (i-\varepsilon, i+1+\varepsilon)$. For any subspace $X \subset \mathbb{R}$, an associated covering for X is $X \cap U$. This kind of covering can be easily generalized into $\mathbb{R}^n$ using the Cartesian products. The parameter ε measures the "resolution" of the covering. A smaller ε will give a larger covering with more open sets. Moreover, the larger covering will reflect finer topological features of the given set in the following sense.

Given a covering U, the nerve associated to U is a simplicial complex defined as follows:

1) The 0-simplices consists of all U_i, i.e., each U_i serves as a vertex in this simplicial complex.
2) A collection of k-vertices $U_{i_1}, U_{i_2}, \ldots, U_{i_k}$ forms a $(k-1)$-simplex if and only if $\cap_{j=1}^{k} U_{i_j} \neq \emptyset$.

The essential steps of the Mapper algorithm can be summarized as follows:

1) Given a data set X, the first step is to select an appropriate continuous function $f: X \to \mathbb{R}^n$. A well-chosen function should effectively capture the essential topological features of the dataset. Commonly used functions include projection function which projects each point onto specific coordinates, and centralization function which computes the average distance of each point to all other points in the dataset.
2) A covering U is chosen for the image $f(X)$. For each open set $U \in U$, the preimage $f^{-1}(U) \subset X$.
3) For each preimage $f^{-1}(U)$ in X, a standard clustering algorithm is applied to separate this preimage into several subsets. Each subset is referred to as a cluster. These clusters together form a covering of the dataset X.
4) The output simplicial complex is the nerve of the covering on X.

The target space $\mathbb{R}^n$ of the filter function can be adapted for specific applications. For a point cloud, for example, height can be used as a lens function. After clustering the points within each preimage, the connectivity of the point cloud can be projected onto the nerve of the resulting cover. Several software packages are available for implementing the Mapper algorithm (Marco-Sola et al., 2012; Veen et al., 2019; Bui et al., 2020). In this work, the Kepler Mapper package is used (Veen et al., 2019). To illustrate the main idea, a simple example is given below before the method is applied to more complex porous-media cases. In practice, the choice of the filter (lens) function and the covering parameters strongly influences the resulting graph. Common filter functions include coordinate projections, distance-based functions, and density-related measures (Singh et al., 2007; Carlsson, 2009). Among these, projection functions are especially useful because they are simple and effective at revealing large-scale geometric patterns in point-cloud data. The range of the filter function is typically covered by overlapping intervals of equal length, which balances topological fidelity against computational cost (Veen et al., 2019; Edelsbrunner and Harer, 2022). In the present work, because the void points of the porous medium are naturally embedded in Euclidean space, coordinate pro-

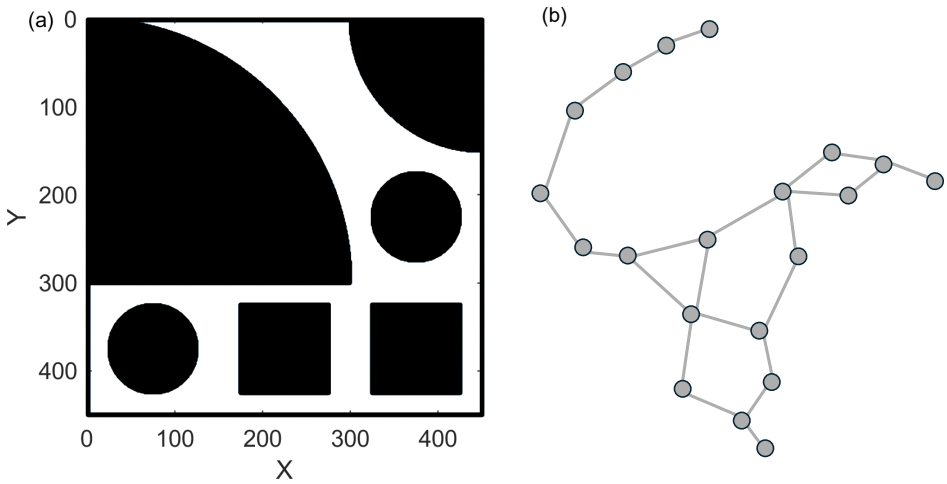

Fig. 2. (a) Example porous model, where black denotes the solid phase and white denotes the void region, and corresponding (b) mapper output graph.

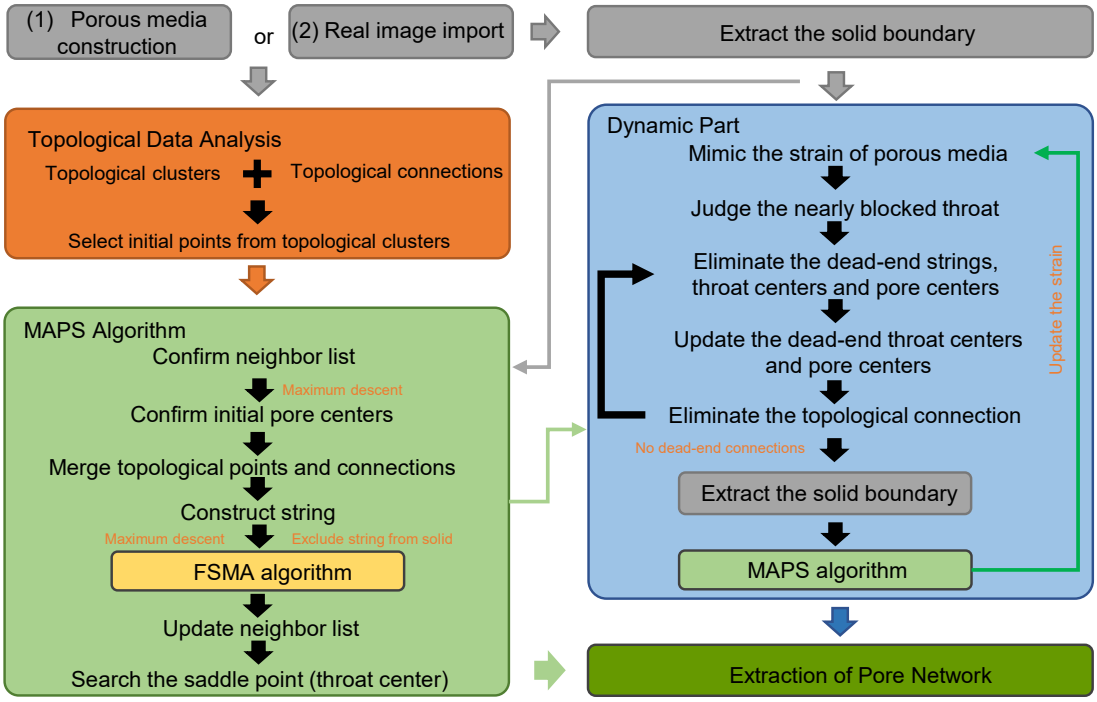

Fig. 3. Workflow of the TPF algorithm.

jections provide a natural choice of filter function, and the cover is therefore constructed from evenly spaced overlapping intervals.

The Mapper algorithm is then used to generate a simple graph that reveals the holes and tubes in the toy image (Fig. 2(a)). The selected lens function is the projection onto the first two coordinates of each point. Because the image of this lens function lies in a 2D space, the cover is constructed as the Cartesian product of interval covers, with each interval divided into 10 equal-length subintervals that overlap adjacent intervals by 10% of their length. The resulting graph is shown in Fig. 2(b). From this graph, the number and relative locations of the holes can be identified clearly.

2.3 Framework of the TPF algorithm

The TDA provides the missing topological link between these two methods. As illustrated in Fig. 3, the porous-medium model is first constructed by image segmentation or artificial generation. The TDA is then applied to delineate the void space and extract topological connectivity. Based on this information, initial strings are generated for the MAPS, and the FSMA is subsequently used to refine these strings toward the medial line of the pore space. The framework also includes a dynamic module to account for structural evolution in porous

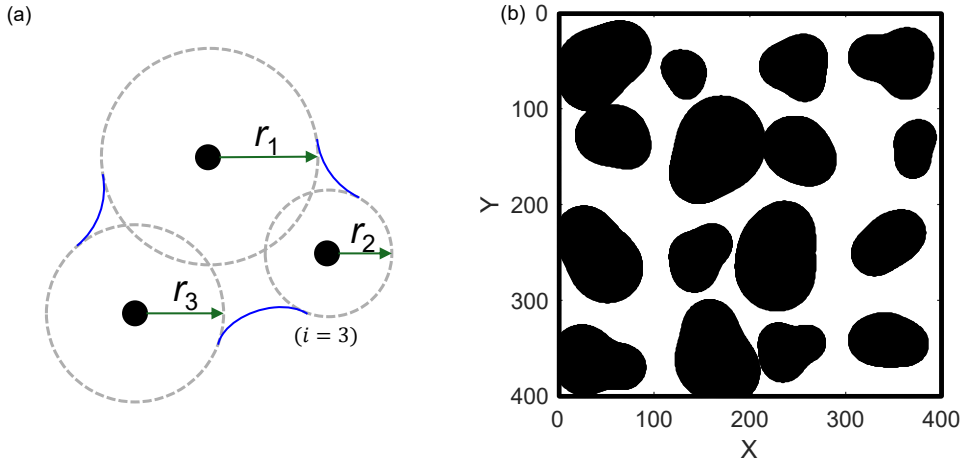

Fig. 4. Construction of the random porous media model: (a) The method to construct the random solid phase and (b) the random porous media model.

rocks.

2.3.1 Porous model construction and initial pore center identification

To construct the porous medium, a set of points is randomly generated within a prescribed domain and assigned corresponding radii. Using the criterion illustrated in Fig. 4(a), the solid phase is then confined within the domain. This procedure yields the random porous model shown in Fig. 4(b), with irregular void and solid regions. Such random and geometrically complex structures provide a useful testbed for evaluating the stability of the algorithm and its sensitivity to challenging pore geometries. In this model, the external boundaries are treated as solid.

Once the porous model has been constructed (Fig. 5(a)), the void phase is extracted (Fig. 5(b)). In Fig. 5(c), the void space is then separated into distinct clusters, after which dead-end clusters that do not contribute to flow are removed (Fig. 5(d)). Each cluster contains many points, from which a subset is randomly sampled to represent the cluster (Fig. 5(e)). Applying the FSMA to these sample points drives them toward pore centers, so that the points converge around each center (Fig. 5(f)). These points are then merged into a single representative point (Fig. 5(g)), which greatly simplifies the connectivity relations and improves computational efficiency. The specific operations are described by Eqs. (6)-(9). Finally, the connectivity between clusters is determined, yielding the required topological connections (Fig. 5(h)) and, consequently, the pore-center information used in the subsequent calculations (Fig. 5(i)).

$$C_j = \left\{ \mathbf{x} \in \mathbf{X} \,\middle|\, \left\| \mathbf{x} - \mathbf{c}_j \right\|_2^2 \le R_c^2 \right\} \tag{6}$$

$$\mathbf{X} = \bigcup_{j=1}^{k} C_j \tag{7}$$

$$\mathbf{c}_j = \frac{1}{|C_j|} \sum_{\mathbf{x} \in C_j} \mathbf{x} \tag{8}$$

$$\mathbf{X}_{\text{new}} = \left\{ \mathbf{c}_j \right\}, j = 1, \ldots, k \tag{9}$$

where $\mathbf{X}$ is the initial dataset; C denotes the data cluster; R_c denotes the cutoff distance to judge the formation of cluster; $|C_j|$ is the number of elements in C_j; k denotes the number of clusters; $\mathbf{c}$ represents the center of the cluster; $\mathbf{X}_{\text{new}}$ means the dataset after merging the initial data.

2.3.2 Optimization of initial string

During construction of the initial string, the complexity of subsurface porous media may cause part of the string to intersect the solid phase, which complicates or even prevents subsequent identification of the medial axis. As illustrated in Fig. 6, the discretization points on the string are progressively displaced until they leave the solid region. For each point, the first direction that moves it out of the solid phase is taken as the optimal correction. This procedure yields a physically admissible connecting path between the two initial pore centers.

The full process from initial-string construction to the final optimized pore-network structure is illustrated in Fig. 7. After the initial points have stabilized at the pore centers, an initial string is constructed (Fig. 7(a)). This string is discretized using a prescribed number of points; in the example shown, six points are used. The procedure in Fig. 6 is then applied to move discretization points out of the solid region, producing the corrected initial string shown in Fig. 7(b). Following the minimum-energy-path idea used in the MAPS (Zhang et al., 2024), the string is then updated toward the medial axis. The pore throat is identified at the local minimum of the DF along each string, after which the pore-network structure can be extracted (Fig. 7(c)).

In pore space, the identified throats are generally curved rather than straight. These throats can be characterized by geometric descriptors such as length, curvature, tortuosity, and minimum cross-sectional radius. In the final pore-network representation, pore centers are treated as nodes and throats as edges carrying these geometric attributes. The extracted net-

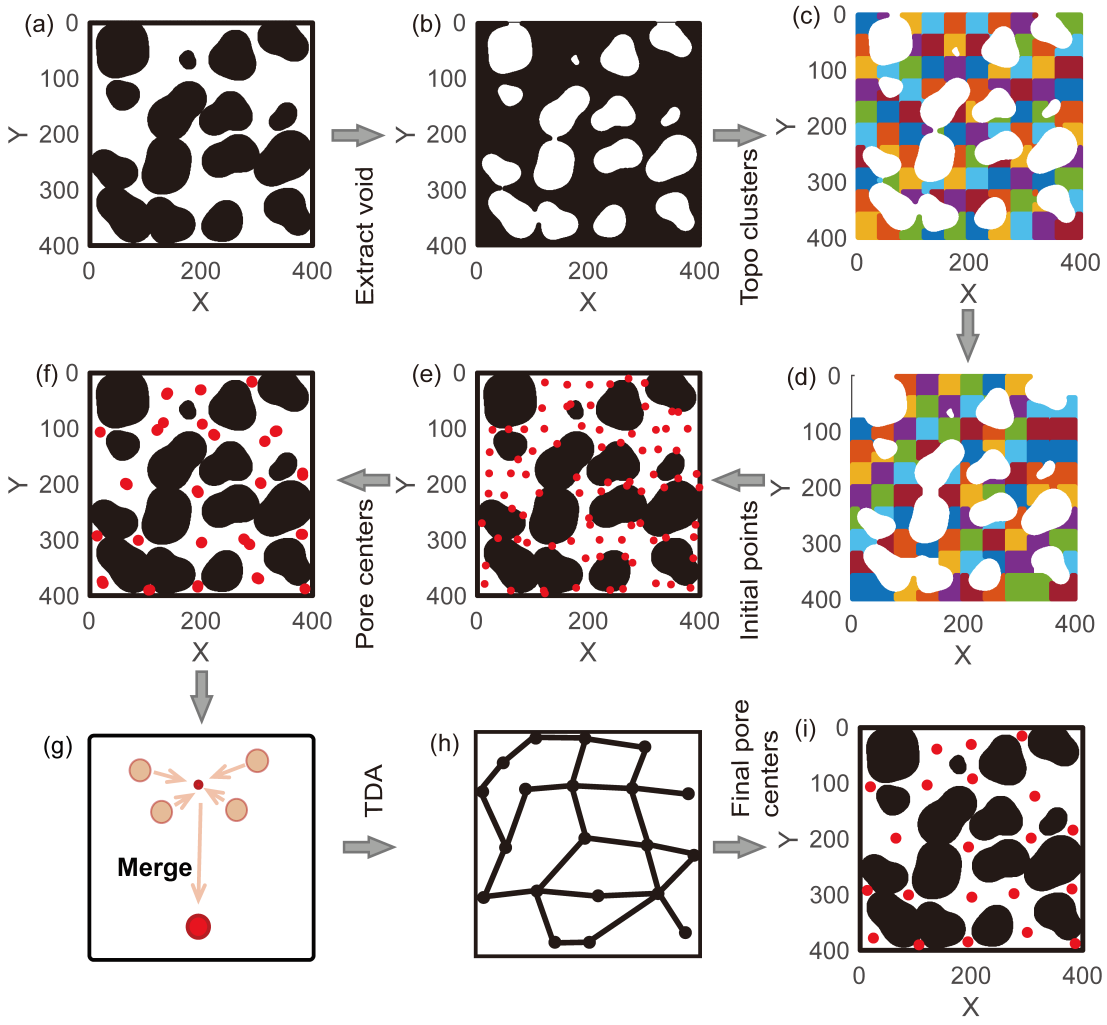

Fig. 5. Identification of the initial pore centers: (a) Porous-medium model, where black indicates the solid phase and white the void phase, (b) inverse model of (a), (c) TDA-based separation of the void space, (d) result after removing dead-end pores, (e) initial sampling points in the void phase, (f) relocation of sampling points toward pore centers, (g) merging each point cluster into a single representative point, (h) topological connectivity of the void phase and (i) final pore-center locations.

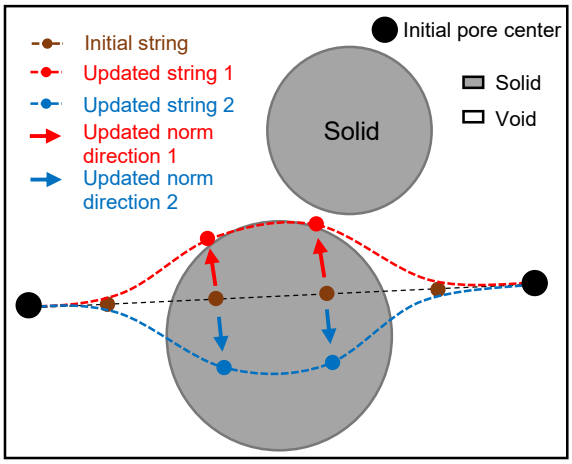

Fig. 6. Schematic illustration of removing the string from the solid phase.

work therefore preserves both the essential topological connectivity and the quantitative geometric information required for subsequent petrophysical analysis.

2.3.3 Manipulations to accelerate calculation

In subsurface environments, rock pore structures evolve continuously, which makes the dynamic PNE desirable. When the porous medium changes, narrow throats may close and block fluid transport. In such cases, as illustrated in Fig. 8, nearly closed channels are removed according to the following procedure:

$$C_T = H\left(D(v,S) - c\right) \tag{10}$$

$$H(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \tag{11}$$

where c is the cutoff distance used to determine the throat state, H is the Heaviside function, x is the function variable, C_T is used to judge whether the throat is closed or open, where $C_T = 0$ means closed and $C_T = 1$ means open. As noted above,

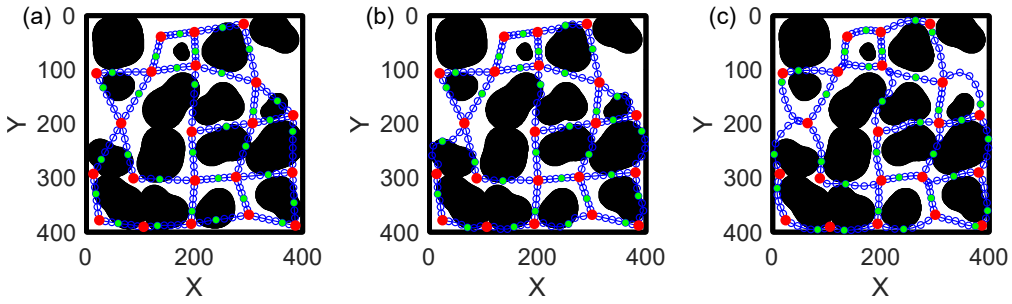

Fig. 7. (a) Initial connections between pore centers, (b) string after removing discretization points from the solid phase and (c) final medial axis.

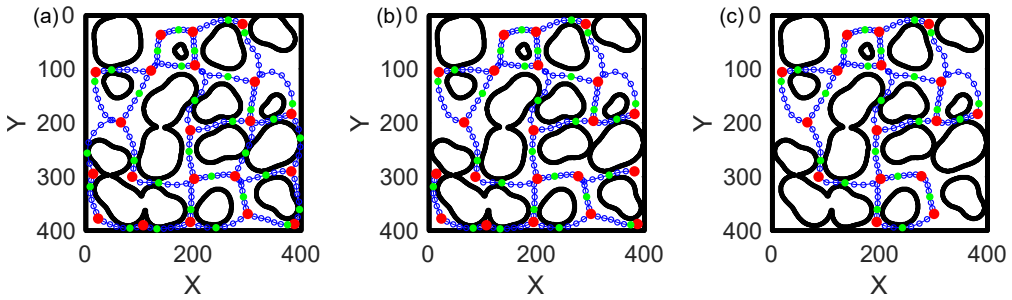

Fig. 8. Elimination procedure for changes in porous media: (a) Pore network without cutting, (b) pore network after cutting and (c) elimination of dead-end strings.

dead-end strings can also be removed to save computational cost.

Computing the DF against the entire solid phase can be expensive, especially for high-resolution models. In most cases, however, only the solid boundary is required. Extracting the boundary therefore substantially reduces the computational burden. This optimization is expected to be even more beneficial for systems with a high solid fraction, such as low-porosity shale.

2.4 Sensitivity and flexibility analysis

This section examines the effect of the number of discretization points used along the string, as shown in Fig. 9. Increasing the number of points improves the accuracy of pore-center and throat identification, but it also raises the computational cost. In practice, an extremely dense string is not required to capture the essential pore-network information. Indeed, the results in Fig. 9(c) and 9(d) are already very similar. An appropriate balance between accuracy and efficiency is therefore needed.

Here, the result obtained with 50 discretization points is taken as the reference solution because of its high accuracy. The spatial error is evaluated using the formula $\text{error} = (C - C_{50})/(\sqrt{L_x^2 + L_y^2})$ and the results are shown in Fig. 10(a). On this basis, eight points provide a practical compromise, yielding an error below 2% while being substantially faster than the cases with 16 or 50 points. Accordingly, the eight-point configuration is adopted in most subsequent tests. The calculation to the solid boundary markedly improves efficiency (Fig. 10(b)).

Three porous models with different resolutions are constructed to assess the stability of TPF, as shown in Fig. 11. All models share the same underlying porous structure, allowing direct comparison. As seen in Fig. 11(a)-11(c), the extracted networks remain stable across resolutions. The lower resolution substantially reduces computational time (Fig. 11(d)). Using only the solid boundary provides additional acceleration, consistent with the previous discussion.

2.5 Application to rock images and quasi-dynamic extraction

The TPF algorithm is next applied to a real rock porous model (Fig. 12). Before extraction, isolated pores are removed because disconnected void space does not contribute to flow. The results show that TPF performs well for pore-network extraction in realistic rock images.

Although the original digital rock images are represented in pixel or voxel form, as in Fig. 12, the proposed TPF algorithm operates in continuous space. Pixels or voxels are used only as the raw input representation of the porous medium. Once the void and solid boundaries have been identified, pore centers and throats are extracted without further reliance on the underlying grid. The final pore network is therefore pixel-free, which distinguishes the TPF from conventional pixel-based skeletonization methods.

A key feature of the TPF is its quasi-dynamic extraction module (Fig. 13), which enables the method to handle structural evolution in porous media. Underground rocks may un-

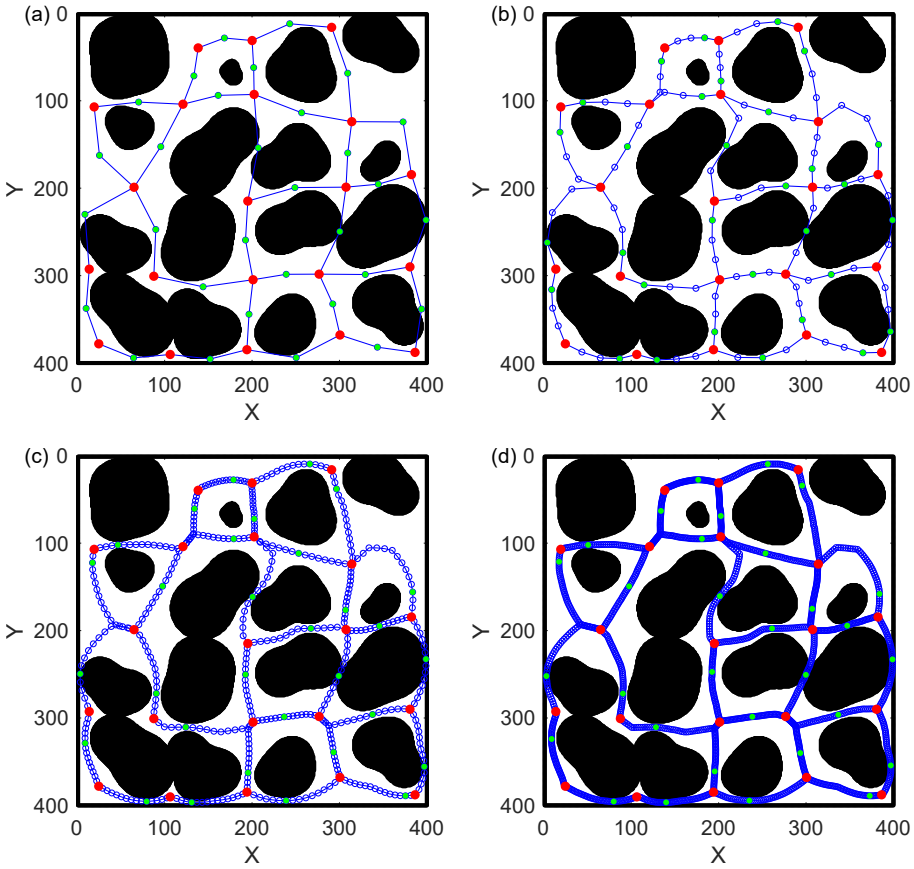

Fig. 9. Extraction results obtained with different numbers of discretization points on the string: (a) 3 points, (b) 5 points, (c) 16 points and (d) 50 points.

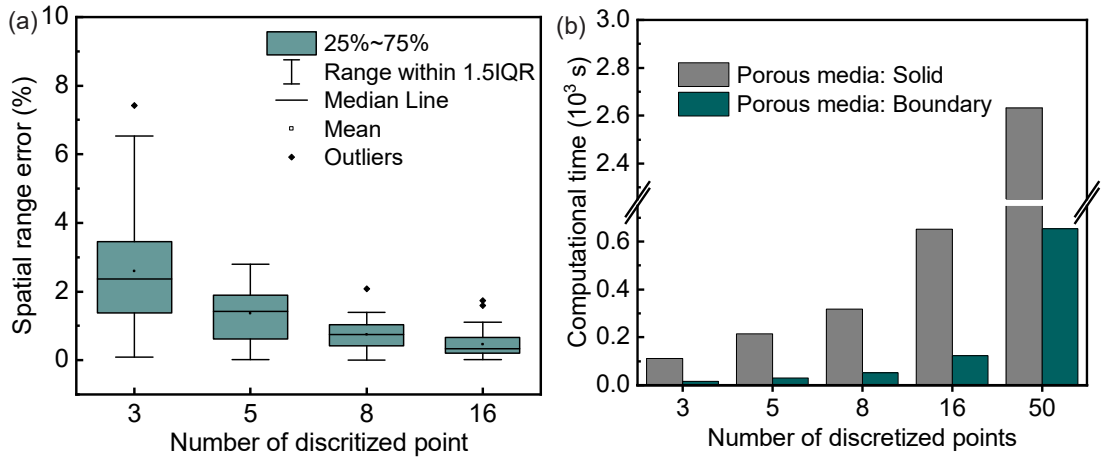

Fig. 10. (a) Spatial error and (b) computational performance for cases with different numbers of discretization points.

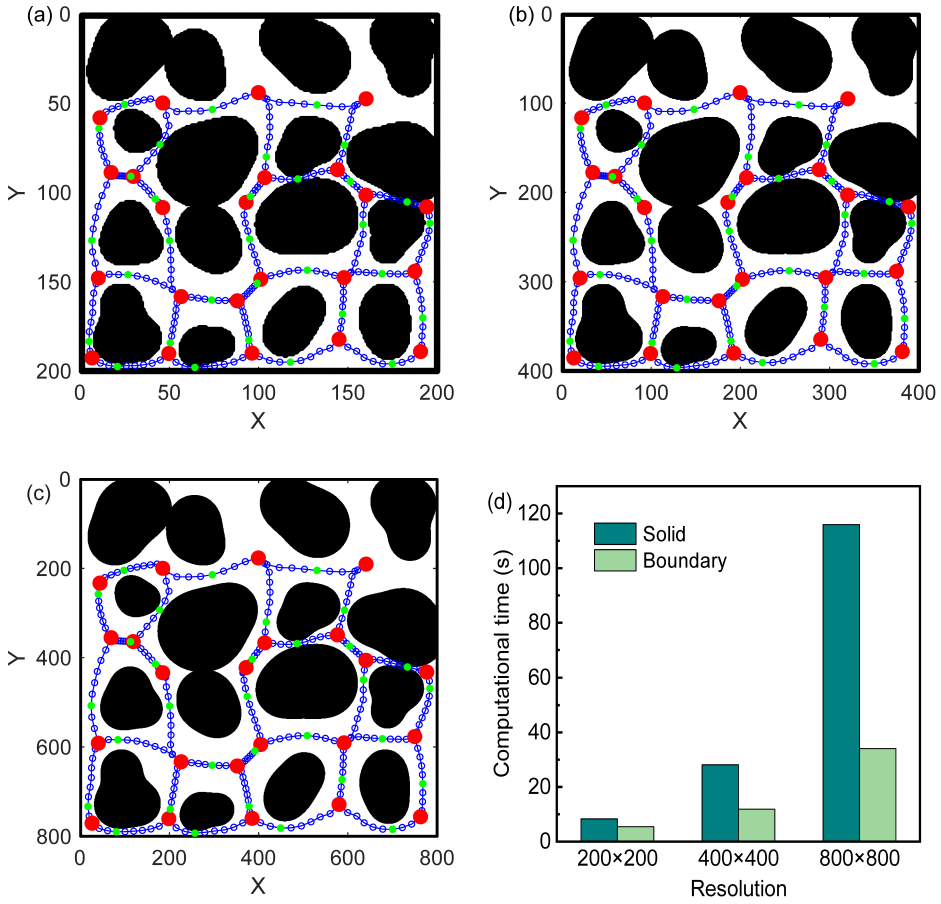

Fig. 11. PNE results for porous models with different resolutions: (a) 200×200, (b) 400×400, (c) 800×800 and (d) computational performance at different resolutions.

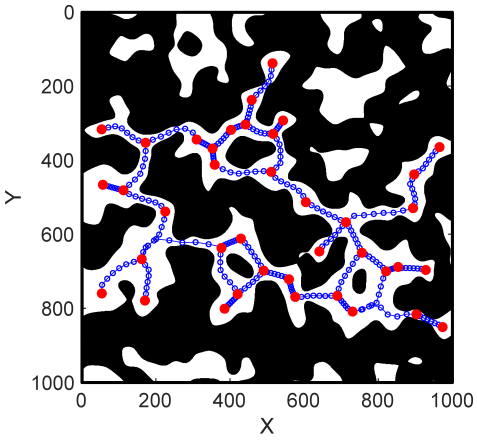

Fig. 12. Pore-network extraction in real binary porous rock model.

dergo compression or swelling as reservoir pressure changes during subsurface operations. In conventional approaches, any structural change requires the pore network to be re-extracted from scratch, which is time consuming. By contrast, the TPF updates the network efficiently by reusing string information from the previous state (Fig. 14). This makes the method much more efficient for tracking pore-network evolution in deforming systems.

To further clarify the advantages of the proposed TPF algorithm, Table 1 compares it with several widely used PNE methods. Each method has distinct strengths and limitations in terms of accuracy, sensitivity to image resolution, computational efficiency, and implementation complexity. The comparison highlights how the TPF combines the key advantages of existing approaches while mitigating several of their main limitations.

Although the main focus of this study is the extraction methodology itself, networks produced by the TPF are fully compatible with existing pore-network flow simulators. The extracted networks provide not only topological connectivity between pore clusters but also the structural basis required for modelling flow and transport processes, including single- and multiphase displacement (Liu et al., 2024a; Liu et al., 2025), reactive transport, and capillary phenomena (Yang et al., 2025). Flow simulations are beyond the scope of the present work, but networks extracted by the TPF can be readily integrated into established modelling frameworks, providing a direct pathway to future applications.

To further validate the method in three dimensions, a 400

Table 1. Advantages and limitations of different PNE approaches.

Method	Advantages	Limitations
Medial Axis (Lindquist et al., 1996)	Preserves topology of pore space. Widely used, conceptually simple.	Produces noisy skeletons in irregular pores (Silin and Patzek, 2006; Zhang et al., 2024). High computational cost in large models.
Classical Maximal Ball	Geometrically intuitive. Provides direct pore and throat sizes.	Strongly resolution-dependent (Silin and Patzek, 2006). Misses small features at low resolution.
FSMA (Liu et al., 2024b)	Works in continuous space. Less resolution-sensitive.	Step-by-step process limits parallelization. Needs pre-information of pore connectivity.
MAPS (Zhang et al., 2024)	Efficient and parallelizable. Robust medial axis identification.	Requires prior connectivity information. Sensitive to initial path selection.
TPF (This study)	Pore connectivity are well preserved. Pixel-free (more accurate). Efficient and less sensitive to image resolution. Supports quasi-dynamic updates under compression/swelling.	Current work focuses on extraction only. Implementation complexity higher than classical methods.

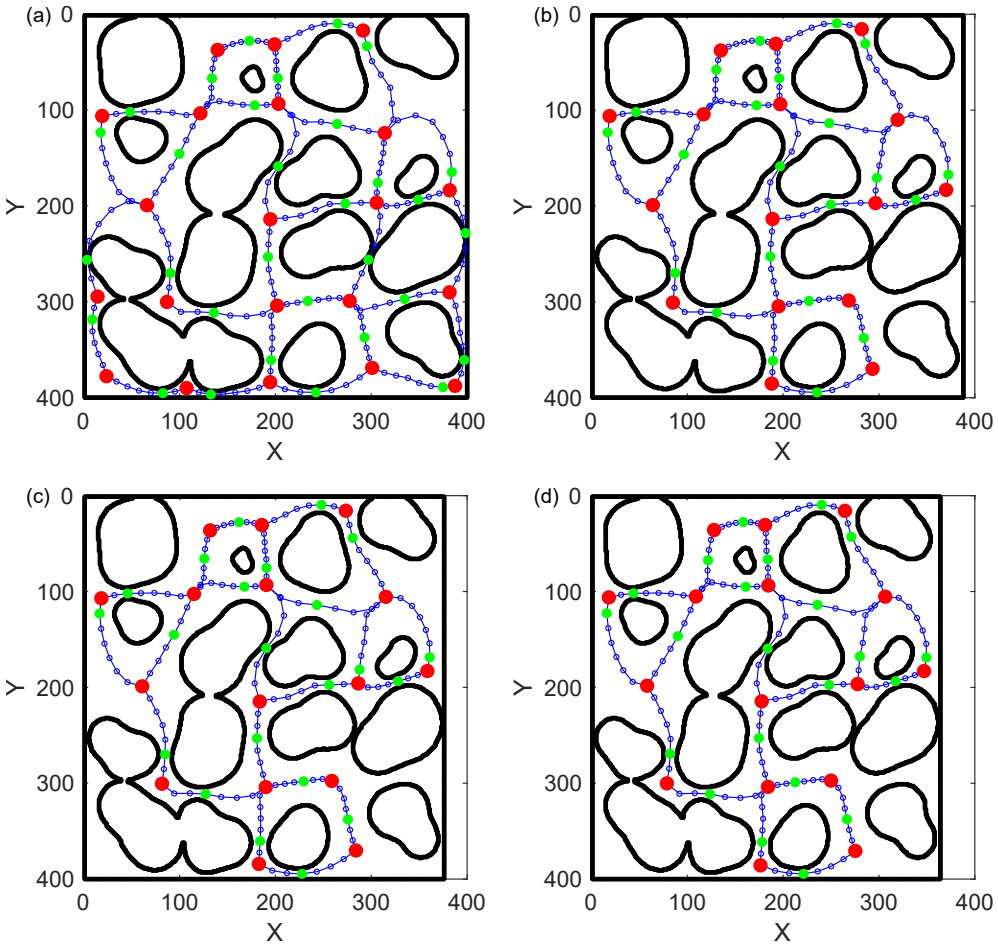

Fig. 13. Quasi-dynamic PNE during compression at strains of (a) 0.0%, (b) 3.0%, (c) 6.0% and (d) 9.0%.

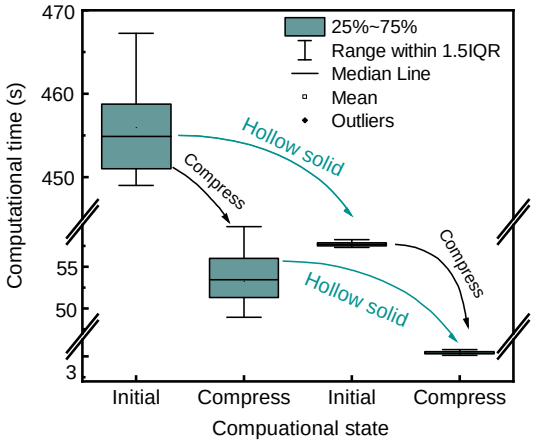

Fig. 14. Computational performance for extraction in the initial and compressed states.

$\times 400 \times 400$-voxel digital rock sample is analyzed. The extracted three-dimensional pore network (Figs. 15(a)-15(c)) confirms that the TPF reconstructs a connected structure suitable for subsequent transport analysis. As shown in Fig. 15(d) and 15(e), the pore-radius and throat-radius distributions obtained by the TPF follow the same overall trends as those from the classical maximal-ball method, indicating that the proposed algorithm preserves the main network-scale geometric statistics in 3D. At the same time, the TPF yields slightly higher frequencies in the small-radius range, suggesting greater sensitivity to fine pores and narrow throats that may be merged or overlooked by the classical method. This difference mainly affects the fine-scale tail of the distributions, while the dominant structural characteristics remain consistent between the two approaches. As shown in Fig. 15(f), the TPF also reduces computational time substantially relative to the classical maximal-ball method, and the acceleration becomes more pronounced when fewer discretization points are used. To assess the practical consequences of the extracted networks, a CO_2 injection simulation is further performed in Fig. 15(g), with methodological details reported elsewhere (Liu et al., 2026). The saturation-injection curves from the TPF-based and maximal-ball-based networks nearly overlap throughout the displacement process, showing highly consistent evolution trends and very similar final saturation levels. This close agreement indicates that the two extracted networks produce essentially the same macroscopic displacement response, further supporting reliability of the TPF for preserving flow-relevant features of the original digital rock.

3. Conclusions

In this study, the TPF algorithm is proposed, which integrates the TDA, MAPS, and FSMA for efficient pore-network extraction in porous media. The main findings are summarized as follows:

1) The TDA provides a reliable description of pore connectivity, allowing the algorithm to distinguish essential links between pore clusters while filtering out topologically irrelevant features.
2) By combining topological dimensionality reduction with the string method, the TPF substantially reduces computational cost while maintaining extraction accuracy.
3) The combination of MAPS-based string evolution and solid-boundary simplification further improves efficiency, making the TPF suitable for large-scale and quasi-dynamic simulations.
4) The TPF produces consistent results across different image resolutions, alleviating one of the main limitations of traditional pixel-based methods.
5) The quasi-dynamic module enables efficient updating of pore networks under compaction or swelling, avoiding full re-extraction from scratch.

Overall, the TPF provides a generalizable framework for pore-network extraction in both synthetic porous models and real rock images. Its topological foundation makes it robust to local noise and resolution effects, while its geometric augmentation preserves quantities needed for comparison with conventional pore-scale descriptors. At the same time, the quantitative petrophysical values derived from the extracted networks will still depend on factors such as rock type, imaging resolution, and preprocessing conditions. Accordingly, although the framework itself is broadly applicable, specific numerical results should be generalized with appropriate caution. Future work will extend the TPF toward more detailed geometric characterization of pore structures and parallel high-performance implementations to further improve both accuracy and efficiency.

Acknowledgements

We would like to express appreciation to the following financial support: Shandong Excellent Young Scientist (Overseas) Program (No. 2024HWYQ-050), National Key Research and Development Project of China (No. 2023YFA1011701), National Natural Scientific Foundation of China (Nos. 12301554 and 12401084).

Conflicts of interest

The authors declare no competing interest.

References

Biasotti, S., Giorgi, D., Spagnuolo, M., et al. Reeb graphs for shape analysis and applications. Theoretical Computer Science, 2008, 392(1-3): 5-22.

Bui, Q.-T., Vo, B., Do, H.-A. N., et al. F-mapper: A fuzzy mapper clustering algorithm. Knowledge-Based Systems, 2020, 189: 105107.

Cai, J., Wei, W., Hu, X., et al. Electrical conductivity models in saturated porous media: A review. Earth-Science Reviews, 2017, 171: 419-433.

Carlsson, G. Topology and data. Bulletin of the American Mathematical Society, 2009, 46(2): 255-308.

Chareyre, B., Cortis, A., Catalano, E., et al. Pore-scale modeling of viscous flow and induced forces in dense sphere packings. Transport in Porous Media, 2012, 94(2): 595-

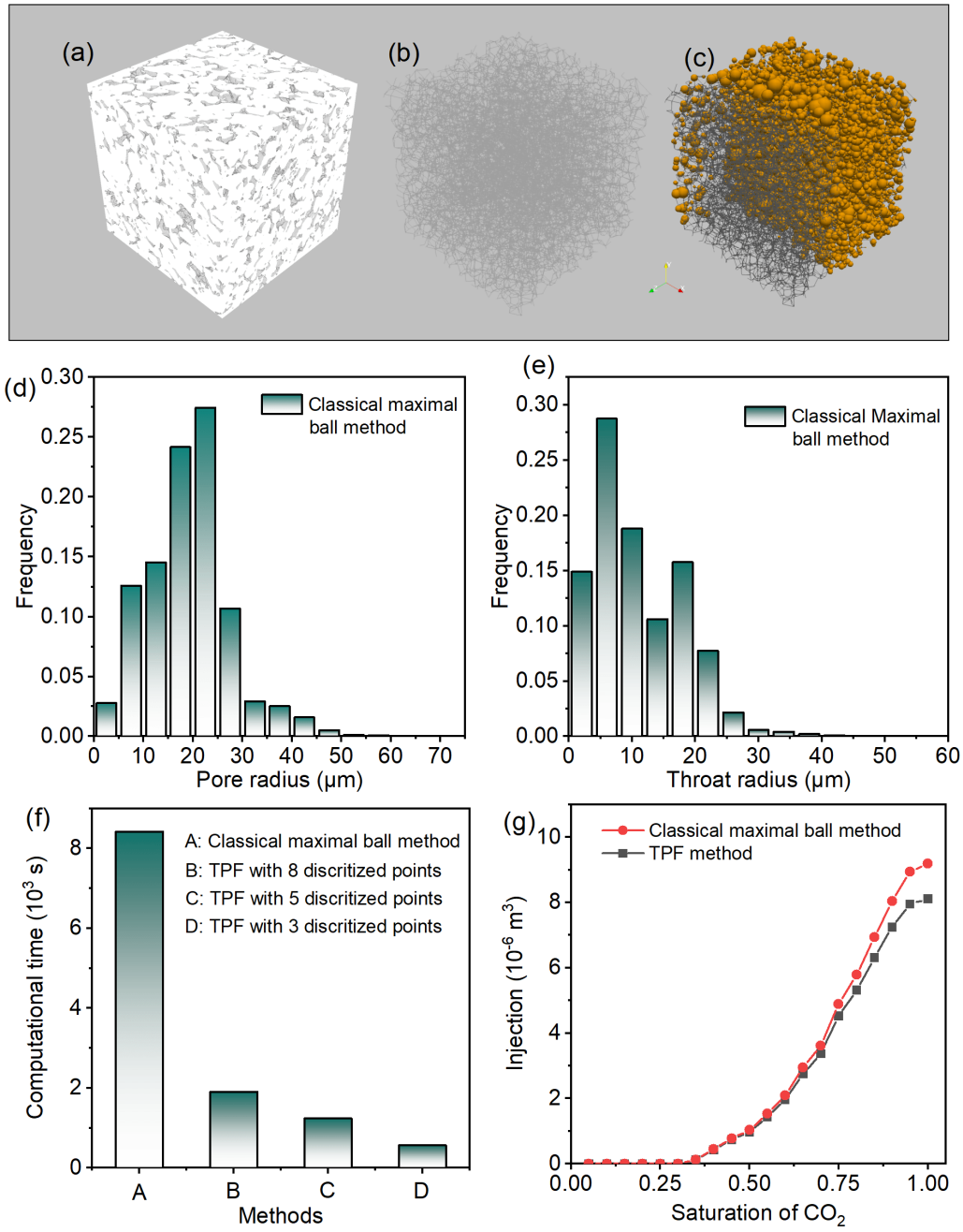

Fig. 15. (a) 3D digital rock model, (b) extracted pore network, (c) CO_2 distribution in the pore network, (d) pore-radius distribution, (e) throat-radius distribution for networks extracted by the classical maximal-ball method and the TPF method; (f) comparison of computational time for different models and (g) CO_2 injection volume versus CO2 saturation.

615.

Chen, S., Li, X., Zhu, G., et al. Classification, controlling factors, and multi-scale characterization techniques in shale reservoir pores: A comprehensive review. Gas Science and Engineering, 2025, 140: 205662.

Cui, R., Hassanizadeh, S. M., Sun, S. Pore-network modeling of flow in shale nanopores: Network structure, flow principles, and computational algorithms. Earth-Science Reviews, 2022, 234: 104203.

E, W., Ren, W., Vanden-Eijnden, E. String method for the study of rare events. Physical Review B, 2002, 66(5): 052301.

Edelsbrunner, H., Harer, J. L. Computational topology: An introduction. Providence, USA, American Mathematical Society, 2022.

Feng, X., Liu, J., Shi, J., et al. Phase equilibrium, thermodynamics, hydrogen-induced effects and the interplay mechanisms in underground hydrogen storage. Computational Energy Science, 2024, 1(1): 46-64.

Jiang, Z., Van Dijke, M. I. J., Geiger, S., et al. Pore network extraction for fractured porous media. Advances in Water Resources, 2017, 107: 280-289.

Lee, T.-C., Kashyap, R. L., Chu, C.-N. Building skeleton models via 3D medial surface axis thinning algorithms. CVGIP: Graphical Models and Image Processing, 1994, 56(6): 462-478.

Lindquist, W. B., Lee, S. M., Coker, D. A., et al. Medial axis analysis of void structure in three-dimensional tomographic images of porous media. Journal of Geophysical Research: Solid Earth, 1996, 101(B4): 8297-8310.

Liu, F., Zhang, Z., Liao, B., et al. Recent advances in phase change microcapsules for oilfield applications. Advances in Geo-Energy Research, 2025, 16(3): 211-228.

Liu, J., Tang, Q., Kou, J., et al. A quantitative study on the approximation error and speed-up of the multi-scale MCMC (Monte Carlo Markov chain) method for molecular dynamics. Journal of Computational Physics, 2022, 469: 111491.

Liu, J., Wang, K., Song, H., et al. Porous media flow modeling from molecular simulations to pore-network modeling: Physics-consistent upscaling of interfacial transport parameters. Computational Geosciences, 2026, 30(2): 15.

Liu, J., Zhang, T., Sun, S. Molecular insights into the carbon dioxide sequestration in kerogen: An accelerated algorithm coupling molecular dynamics simulations and Monte Carlo methods. Process Safety and Environmental Protection, 2024a, 185: 1336-1351.

Liu, J., Zhang, T., Sun, S. A new pixel-free algorithm of pore-network extraction for fluid flow in porous media: Flashlight search medial axis. Advances in Geo-Energy Research, 2024b, 13(1): 32-41.

Marco-Sola, S., Sammeth, M., Guigó, R., et al. The GEM mapper: Fast, accurate and versatile alignment by filtration. Nature Methods, 2012, 9(12): 1185-1188.

Morimoto, T., Zhao, B., Taborda, D. M., et al. Critical appraisal of pore network models to simulate fluid flow through assemblies of spherical particles. Computers and Geotechnics, 2022, 150: 104900.

Ni, X., Chen, W., Li, Z., et al. Reconstruction of different scales of pore-fractures network of coal reservoir and its permeability prediction with Monte Carlo method. International Journal of Mining Science and Technology, 2017, 27(4): 693-699.

Niasar, V., Hassanizadeh, S., Pyrak-Nolte, L., et al. Simulating drainage and imbibition experiments in a high-porosity micromodel using an unstructured pore network model. Water Resources Research, 2009, 45(2): W02430.

Qin, X., Wang, H., Xia, Y., et al. Micro-and nanoscale flow mechanisms in porous rocks based on pore-scale modeling. Capillarity, 2024, 13(1): 24-28.

Raoof, A., Hassanizadeh, S. M. A new method for generating pore-network models of porous media. Transport in Porous Media, 2010, 81: 391-407.

Shao, J., You, L., Jia, N., et al. Investigation of induced change in pore structure by the reaction of shale with fracturing fluid. Gas Science and Engineering, 2023, 110: 204860.

Silin, D., Patzek, T. Pore space morphology analysis using maximal inscribed spheres. Physica A: Statistical Mechanics and Its Applications, 2006, 371(2): 336-360.

Singh, G., Mémoli, F., Carlsson, G. E. Topological methods for the analysis of high dimensional data sets and 3D object recognition. Eurographics Symposium on Point-Based Graphics, 2007, 2: 91-100.

Thompson, K. E., Fogler, H. S. Modeling flow in disordered packed beds from pore-scale fluid mechanics. AIChE Journal, 1997, 43(6): 1377-1389.

Veen, H., Saul, N., Eargle, D., et al. Kepler Mapper: A flexible Python implementation of the Mapper algorithm. Journal of Open Source Software, 2019, 4(42): 1315.

Yang, L., Liu, Z., Zhao, Z., et al. Experimental study of carbonated water imbibition in deep coal rocks using nuclear magnetic resonance spectroscopy. Capillarity, 2025, 16(2): 27-38.

Zhang, T., Salama, A., Sun, S., et al. Pore network modeling of drainage process in patterned porous media: A quasi-static study. Journal of Computational Science, 2015, 9: 64-69.

Zhang, Y., Liu, J., Zhang, T., et al. Medial Access Path Search (MAPS) for pore-network extraction. Computational Geosciences, 2024, 28: 979-989.

Zhang, Y., Yang, X., Zhang, L., et al. Energy landscape analysis for two-phase multi-component NVT flash systems by using ETD type high-index saddle dynamics. Journal of Computational Physics, 2023, 477: 111916.