OPEN

Preferential flow pathways in a deforming granular material: self-organization into functional groups for optimized global transport

Joost H. van der Linden [1], Antoinette Tordesillas[2]* & Guillermo A. Narsilio[1]

Existing definitions of where and why preferential flow in porous media occurs, or will occur, assume a priori knowledge of the fluid flow and do not fully account for the connectivity of available flow paths in the system. Here we propose a method for identifying preferential pathways through a flow network, given its topology and finite link capacities. Using data from a deforming granular medium, we show that the preferential pathways form a set of percolating pathways that is optimized for global transport of interstitial pore fluid in alignment with the applied pressure gradient. Two functional subgroups emerge. The primary subgroup comprises the main arterial paths that transmit the greatest flow through shortest possible routes. The secondary subgroup comprises inter- and intra-connecting bridges that connect the primary paths, provide alternative flow routes, and distribute flow through the system to maximize throughput. We examine the multiscale relationship between functionality and subgroup structure as the sample dilates in the lead up to the failure regime where the global volume then remains constant. Preferential flow pathways chain together large, well-connected pores, reminiscent of force chain structures that transmit the majority of the load in the solid grain phase.

Where preferential fluid flow paths in porous media occur, what fraction of flow these paths carry, and why certain pathways are preferred has been defined in various ways, including (1) the delineation of the liquid phase in relatively mobile (preferential) and immobile parts[1], (2) 'all phenomena where water and solutes move along certain pathways, while bypassing a fraction of the porous matrix'[2], [p ∼ 150], and (3) non-equilibrium transport through large interaggregate voids referred to as macropores in soil science[3]. A dedicated body of research has also presented different ways to characterize preferential flow, including dye staining experiments[4,5], mathematical indicators based on the breakthrough curves and early arrival times[6], thresholding the highest flow rates ('hotspots') in numerical simulations (e.g.[7,8]) and, more recently, complex network approaches focusing on individual paths of least resistance[9–12]. While these studies have continuously improved our understanding of preferential flow, predicting emergent preferential flow remains particularly challenging due to the multi-scale, coupled interplay between complex, heterogeneous pore geometry[13], capillary, gravity and viscous forces acting on the fluid[14], pore pressure[15], antecedent moisture[16], organic matter[14] and wettablity effects[17]. Consequently, the emergence of preferential flow is one of the main contributors to uncertainty in the prediction of transport[1,18]. Further progress is impeded by the lack of a quantitative, multi-scale and predictive definition of the structure and location of preferential flow pathways. Indeed, the aforementioned definition (1) and characterization techniques (dye tracing, breakthrough curves and hotspots) cannot predict preferential flow structure without conducting a fluid flow experiment or numerical simulation, while definition (3) does not directly account for the connectivity that is inherent to the chains of pores that make up preferential flow pathways[19].

Where and why preferential flow paths occur is inherently related to the connectivity and internal properties. Such knowledge plays an important role in a wide range of applications, such as soil contamination[3,13], fluid homeostasis in the central nervous system[20], ore dewatering processes[8], landfill waste leaching[21] and volcanic lava flow[22]. The size and properties of the fast-flowing, preferential fraction of the pore network within soils, in particular, have repeatedly been highlighted as an important topic of research that has not been conclusively

[1]Department of Infrastructure Engineering, The University of Melbourne, Melbourne, Australia. [2]School of Mathematics and Statistics & School of Earth Sciences, The University of Melbourne, Melbourne, Australia. *email: atordesi@unimelb.edu.au

1

quantified[13,14,23]. Experimental observations point to a complex network of inter- and intra-connected preferential flow segments with varying size and connectivity, rather than disjoint pathways spanning the entire medium, both at the hillslope scale[16] and in soil columns analyzed using X-ray computerized tomography[19]. Allaire *et al.*[4] conclude that lateral preferential flow is at least as important as vertical preferential flow (in the direction of gravity) and their interaction is critical for the overall throughput, particularly at large scales, yet these connections are rarely studied. Interconnections between preferential flow paths are also known to be a critical factor for the hydraulic properties of aquifers[24], and may increase permeability in fractured media with low matrix permeability and high fracture permeability[25]. Connectivity is generally considered one of the most important influencing factors of permeability in porous media[26–31]. Overall, these observations suggest two natural groups within preferential flow paths, namely primary, highway-like preferential route segments, inter- and intra-connected by secondary, bridge-like pathways that support the overall preferential throughput.

Elucidating the functional preferential flow groups requires a mathematical framework that can account for geometrical and material heterogeneity germane to complex porous media. In this context, *network flow theory* and, in particular, the Maximum Flow, Minimum Cost (MFMC) algorithm, has shown promise in modeling (preferential) force transmission[32–34] in cohesionless and cohesive granular materials in 3D[35,36]. Tordesillas *et al.*[35] observe a 76–82% capture rate of grains predicted as being 'preferential' force transmission pathways by MFMC, measured as a percentage of grains in the (known preferential) force chains (see[37,38]). A similar approach is followed to predict the macro-crack pattern in heterogeneous and multiphase concrete specimens, prior to initiation of actual damage[39]. The use of a network flow model for fluid transmission also has precedent, with Ushizima *et al.*[40] using the maximum flow component of MFMC to characterize fluid storage after precipitation for carbon sequestration applications. Ju *et al.*[12] compare shortest paths, used in the minimum cost component of MFMC, to (experimentally) measured preferential flow in two dimensions and observe good agreement in the primary flow pathway structure. Shortest paths are also employed by Rizzo and Barros[11] to quantify solute transport through heterogeneous porous media, establishing a relationship with the (preferential flow-induced) early arrival times of a solute plume. Further characterization of the optimal flow pathways predicted using a network flow model can be obtained using *complex network theory*. Gao and co-authors, for example, uncover the structure and flow patterns in multi-phase oil-water slug flow using a range of complex network techniques, including PageRank versatility and community structure[41–43]. Given their ability to accurately quantify multi-scale connectivity, analyzing complex systems using complex networks is an increasingly prominent paradigm in numerous scientific areas[44–46], including the geosciences[31,47].

In this work, we propose a definition for preferential flow pathways through porous media, as given by the output of the Maximum Flow, Minimum Cost algorithm. We build on the insights gained from applying network flow models to force transmission by applying the algorithm to the network of pores and throats, capturing the domains in which the bulk of fluid flow occurs. MFMC returns a percolating set of optimal pathways through the medium, pushing 'flow' in the direction of the pressure gradient. We incorporate both local knowledge of transmission (e.g. Hagen-Poiseuille conductance) and the multi-scale connectivity that is inherent to flow pathways, moving beyond the single-scale views associated with existing definitions. Further building on and formalizing the intuitions behind micropore-macropore and mobile-immobile delineations of the void phase, we employ a statistical model to subdivide preferential pathways into the primary routes and the secondary pathways that connect them. MFMC is applied to a fully saturated deforming granular material, to capture part of the dynamics that are inherent to the manifestation of flow in real (fractured) sands, soils and rocks[48,49]. For a validation of the proposed network flow model on a collection of well-known, static sphere assemblies, we refer to the Supplementary Material. Our physics-based and data-driven model is not intended to replace experimental approaches, such as preferential solute transport imaging using high-resolution X-ray tomography[19,50], or micromechanical models coupling fluid flow and mechanical deformation[15]. Rather, our network flow model aims to complement such investigations to gain supplemental insights and further our understanding of where, why and to what degree preferential flow paths form in granular, porous media.

Results

Figure 1 provides a summary of the data generated, methodology used and our key results, further expanded on in this section.

Experiment. To capture the evolution of the pore space in a deforming granular material in the presence of dilatant shear zones, we subject a dry assembly of 6,776 polydisperse spheres to a triaxial shear test. This sample is one of a series of shear tests which has been comprehensively examined and reported elsewhere[10,48,49,51] (see also Ord *et al.*[52,53] for similar tests). The mechanical response of the sample is summarized in Fig. 2. During the test, the assembly expands globally in the horizontal directions under a relatively high constant confining pressure and a constant vertical axial strain rate, as evidenced by the increase in porosity and volumetric strain. Pore bodies increase in size accordingly, as shown in Fig. 2(b), merging smaller pores (decreasing their frequency) and developing a long tail in the distribution. Peak stress ratio occurs at about 7% axial strain and the assembly is in steady state at 20% and beyond. Snapshots of the assembly at 430 regularly-spaced strain intervals between 0 and 30% axial strain are the main input data for the remainder of the analysis.

Maximum flow, minimum cost. A pore network is constructed independently for each snapshot of the deforming assembly, embedding local connectivity of the pore space by modeling pores as nodes and throats as edges in a graph. This particular pore network construction approach has been previously validated in[54,55]. Each pore network is the input for the Maximum Flow, Minimum Cost (MFMC) algorithm. MFMC aims to identify the optimal flow pathways through the pore network, starting at a collection of defined inlet nodes at the top of the assembly (shown in Fig. 1) and ending at the outlet nodes at the bottom. Optimality in this network flow
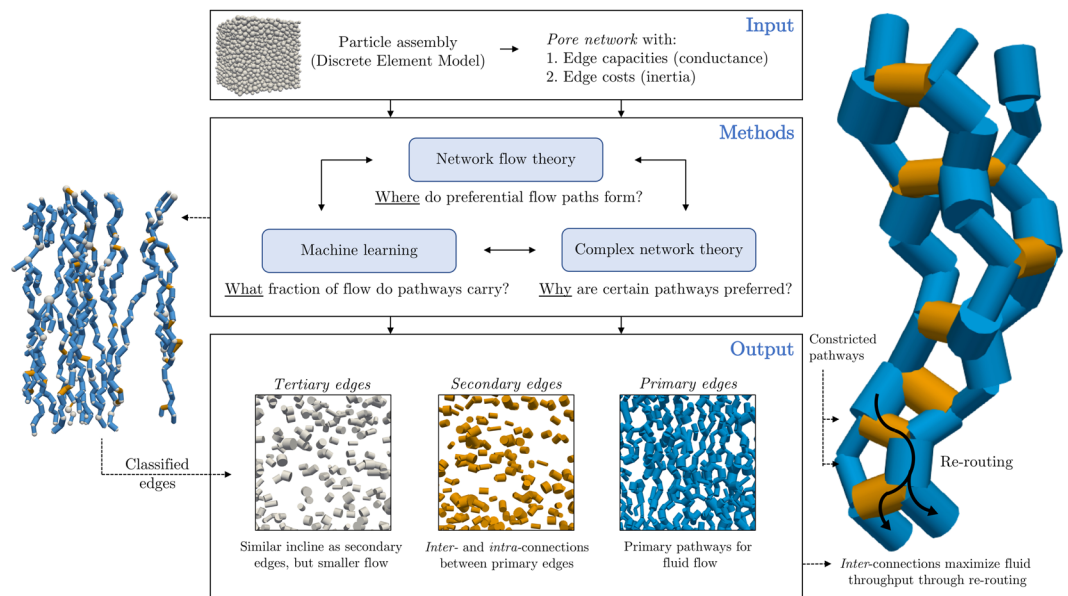
Figure 1. Optimal pathways through a failing granular material give rise to a hierarchy of primary, preferential flow pathways, laterally supported by secondary, bridging pathways. Snapshots of an assembly of spherical particles subjected to a triaxial test are extracted from the discrete element simulation (Input). A pore network is constructed for each assembly, in which nodes represent pores and edges represent throats (Input). Next, the Maximum Flow, Minimum Cost (MFMC) algorithm is applied to the pore network to extract a sub-network P of optimal flow paths. Using a logistic regression classifier, we distinguish between edges in P that can be correctly predicted as being part of P (primary edges, blue), incorrectly predicted as not being part of P (secondary edges, orange) and the remainder (tertiary edges, grey). Further examination of these edges using complex network theory reveals secondary edges act either as intraconnecting pathways, embedded in chains of primary edges, or as interconnections, bridging the primary pathways and re-routing flow to maximize throughput, as shown on the right-hand side.

model is achieved by maximizing *flow* (subject to the *capacity*) while minimizing *cost*. To mimic fluid flow, we define capacity as the fluid conductance of a simplified cylindrical throat, calculated using the Hagen-Poiseuille law. Under the corresponding assumptions of tube flow, wide throats that connect large pores over a short distance have large capacity. Cost is defined as an inertia penalty value. If an edge (i, j) is poorly aligned with other edges connected to nodes i and j, then the cost value is high. This penalty models the resistance of fluid to flow 'around a bend', due to inertia of the fluid. With capacity defined as conductance and cost representing inertia effects, we essentially provide the network flow model with some of the key ingredients of the Navier-Stokes equation for viscous fluid flow. Defining capacity as conductance also allows us to conveniently (and efficiently) compute the approximate numerical permeability. Assuming saturation, we solve the Stokes equation for each pore network, imposing a pressure difference of 1 Pa over the inlet- and outlet nodes and inducing local hydraulic gradients in the order of 10 Pa/m. From the resulting pressures and flow rates, the permeability is derived using Darcy's law. The result is included in Fig. 3(b) and shows that the porosity in Fig. 2(a) and volumetric strain (capturing dilation) in Fig. 2(c) follow the same trend as permeability, as expected. The permeability increases by approximately a factor two between the minimum at 5% axial strain and the maximum in the steady state.

Given the edge capacities and costs, the MFMC algorithm returns a sub-graph (denoted P and referred to as the *preferential flow network*) of nodes and edges that make up the (optimal) *preferential flow pathways* through the pore network. The percolating network P transmits the maximum flow from inlet nodes to outlet nodes, where the maximum flow is the maximum number of units of flow that can be 'pushed' through the network, bounded by edge capacities. Due to the absence of any global physical bottlenecks, the majority (between 72% and 77%) of edges in the original pore networks at different strain intervals are part of the corresponding MFMC sub-graph P. A small excerpt of P is shown on the right-hand side of Fig. 1. The maximum flow and corresponding total (summed) cost of transporting the maximum flow is plotted in Fig. 3. Maximum flow follows approximately the same trend as permeability, suggesting good resemblance between *macro-scale* MFMC and Stokes flow behavior. In the Supplementary Material, the Stokes flow permeability is shown to correspond well to the permeability derived from the Navier-Stokes equation for fully saturated sphere packing geometries. The Navier-Stokes permeability has previously been validated against experimental values[56], suggesting indirectly that MFMC can capture realistic physical fluid flow behavior. Total cost follows the inverse trend of permeability, capturing the intuition of increased permeability in the presence of 'cheaper' flow pathways. At the micro (edge) scale, shown in Fig. 3(a), MFMC flow is well-correlated with Stokes flow rates, with a Pearson correlation coefficient of 0.8 at 20% strain. The (statistically significant) correlation coefficient otherwise varies between 0.72 at 0% strain and 0.81 in the steady state ($p < 2 \cdot 10^{-16}$).
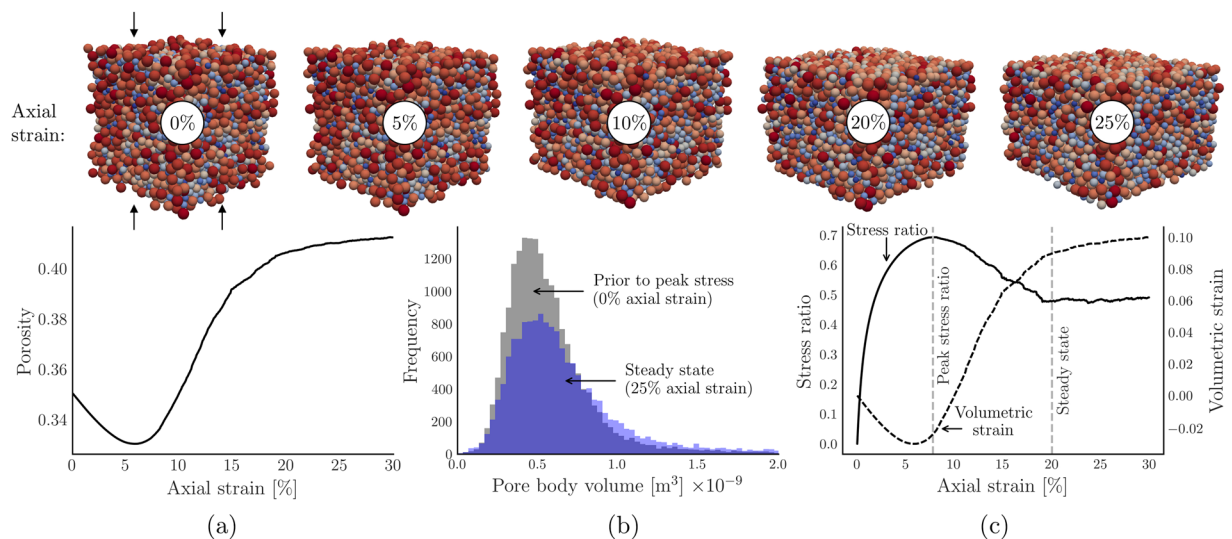
Figure 2. A previously published and validated discrete element simulation is used to model a deforming granular material. Summarized results for the discrete element simulation. Showing (**a**) porosity versus axial strain, (**b**) pore body volume distributions, prior to peak stress and in the steady state, and (**c**) the evolution of the stress ratio and volumetric strain.
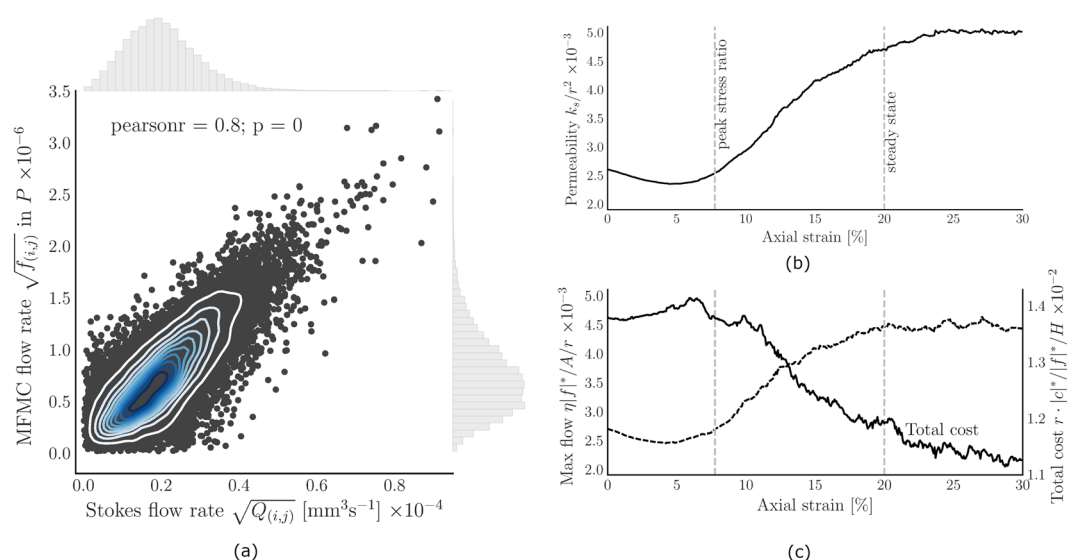

Figure 3. Micro-scale and macro-scale output of the network flow model, in the form of the edge flows, maximum flow and associated cost of transmission shows good agreement with well-known trends in stress ratio, volumetric strain, Stokes flow and permeability for deforming granular materials. (**a**) Correlation between (square-root transformed) network model flow rates and Stokes flow rates at the edge level for 20% axial strain, including marginal distributions and density contours to indicate the joint distribution. (**b**) Evolution of the (dimensionless) permeability, normalized by the squared particle radius r^2, plotted against axial strain as the sample fails. (**c**) Dimensionless maximum flow $|f|^*$ and total cost $|c|^*$, as computed using the Maximum Flow, Minimum Cost algorithm. The dimensionless maximum flow is calculated by multiplying with the dynamic viscosity η and dividing by the cross-sectional area A and mean particle radius r. Total cost is normalized by the mean particle radius r, maximum flow $|f|^*$ and height H of the assembly.

Primary, secondary and tertiary flow.   Having defined our network flow model, and having validated the output of the algorithm by comparing flow at the micro-structural and macro-structural level, we focus on *why* preferential flow occurs in certain edges next. To do so, we define a binary classifier to predict inclusion or exclusion $\chi_{(i,j)}$ of a particular edge in the MFMC sub-graph P, i.e. $\chi_{(i,j)} = 1$ if $(i, j) \in P$ and $\chi_{(i,j)} = 0$ otherwise. For the results in Table 1, the classifier model is fitted on a randomly selected 80% of edges in the sample with axial strain 21.54% (21737 edges). Generalization performance is obtained by testing the fitted model on the remaining 20% of edges, for which we obtain an area under the *ROC* curve equal to 0.76 (0.5 corresponds to random guessing,

Feature name	Equation	Significance	Odds ratio	95% conf. interval
Edge cost	$X_1 = c_{(i,j)}$ in Eq. (4)	$p < 2 \cdot 10^{-16}$	$e^{\beta_1} = 0.55$	[0.53, 0.57]
Throat area	$X_2 = A_{(i,j)}$	$p < 2 \cdot 10^{-16}$	$e^{\beta_2} = 0.84$	[0.81, 0.87]
Adjacent pore volume	$X_3 = \overline{V}_{(i,j)}$ in Eq. (1)	$p < 2 \cdot 10^{-16}$	$e^{\beta_3} = 1.36$	[1.30, 1.42]
Incline	$X_4 = \theta_{(i,j)}$ in Eq. (2)	$p < 2 \cdot 10^{-16}$	$e^{\beta_4} = 2.59$	[2.48, 2.69]
(Constant)		$p < 2 \cdot 10^{-16}$	$e^{\beta_0} = 3.66$	[3.52, 3.80]

Table 1. Features for each edge (i, j) used in the logistic regression model. Notation of features and coefficients refers to Eq. (7).

Figure 4. Confusion matrix for the prediction on the assembly with axial strain 22.94%.

while 1.0 would imply a perfect prediction). For other strain rates (repeating the model fitting process described here), the area under the ROC curve is approximately similar, varying between 0.75 and 0.81.

The odds ratio of the cost value $c_{(i,j)}$ (0.55) in Table 1 is smaller than one, indicating that every unit increase in the standardized value of $c_{(i,j)}$ decreases the odds of $\chi_{(i,j)} = 1$ by a factor 0.55. The odds ratio for throat area is closer to one, suggesting that a larger throat area has a slightly negative impact on the odds of an edge being included in P. For $\overline{V}_{(i,j)}$ and $\theta_{(i,j)}$, we observe $e^{\beta_4} > e^{\beta_3} > 1$, which implies that these features positively impact the odds of being included in P. A unit increase in the standardized incline, in particular, increases the odds by a factor 2.59, capturing the intuition that edges better aligned with the vertical pressure gradient of the assembly are more likely to be part of the preferential flow network. In contrast with throat area, pore volume also has a positive effect on the odds. For a relatively narrow throat size distribution and in the absence of major bottlenecks, we attribute this difference to the preference of flow to find the most direct routes across the sample, as provided by pores of larger size and edges aligned with the global pressure gradient.

Having fitted the logistic regression model, and having tested the ability of the model to generalize, we re-fit the model to 100% of edges in the the assembly with axial strain 21.54% and predict inclusion or exclusion in P for each edge in the pore network at the next strain state (21.61%). The confusion matrix for this result is shown in Fig. 4.

For this sample, 69% of edges in P are correctly predicted as $\chi_{(i,j)} = 1$ (true positives). The remaining 31% are incorrectly predicted as $\chi_{(i,j)} = 0$ (false negatives). The difference between true positives and false negatives gives rise to a distinction between edges in P that can be predicted with physical features and a binary classifier, versus edges in P that are less predictable. As such, we designate edges in P that are correctly predicted as $\chi_{(i,j)} = 1$ as *primary edges*. Edges that are part of P but incorrectly predicted as $\chi_{(i,j)} = 0$ are called *secondary edges*. All edges not in P are referred to as *tertiary edges*. Figure 1 shows an example of this hierarchy. For other strain rates (repeating the model fitting process as described), the true positive rate and true negative rate are approximately similar to the rates shown in Fig. 4, varying between 68% and 75%.

To assess this distinction further, the distribution of the incline of primary, secondary and tertiary flow edges is shown in Fig. 5 along with the Stokes flow rates in corresponding edges.

Clearly, edges closer aligned with the pressure gradient (at 90°) have higher flow rates, as expected. The majority of edges are primary, with an incline centered around approximately 45° and exhibiting high flow rates ($Q_{(i,j)} = 5.9 \cdot 10^{-10}$ m^3 s^{-1}, on average). In this sense, primary edges bear resemblance to the preferential 'hotspots' of fluid flow, as described in the introduction[7,8]. Tertiary edges, in contrast, are more horizontally aligned (21°, on average), perpendicular to the pressure gradient, and have a considerably lower average flow rate ($2.1 \cdot 10^{-10}$ m^3 s^{-1},). Interestingly, secondary edges have even smaller inclines (15°, on average), yet the flow rates are considerably higher ($4.0 \cdot 10^{-10}$ m^3 s^{-1}, on average) than the flow rates of tertiary edges. Indeed, comparing the square-root transformed flow rate for primary, secondary and tertiary edges in a one-way ANOVA test rejects the null hypothesis that the means are the same ($F(2,33963) = 3365$, $p < 2 \cdot 10^{-16}$). A post-hoc Tukey test also confirms that there is a statistically significant difference between each of the three pairs of means ($p < 2 \cdot 10^{-16}$). In contrast, the average pore volume of adjacent nodes is $1.36 \cdot 10^{-9}$ for secondary edges and $1.40 \cdot 10^{-9}$ for tertiary edges, which is not a statistically significant difference (Tukey test, $p = 0.62$, using square-root transformation). In conclusion, secondary edges have a less favorable incline than tertiary edges, and similar adjacent pore volume, yet the flow rates for secondary edges are significantly higher. These conclusions also hold for flow networks at other axial strain states, including pre-peak stress.

These observations suggest that secondary edges play a distinct role in the transmission of fluid through the pore space. Despite less favorable circumstances in terms of incline, secondary edges have significantly higher
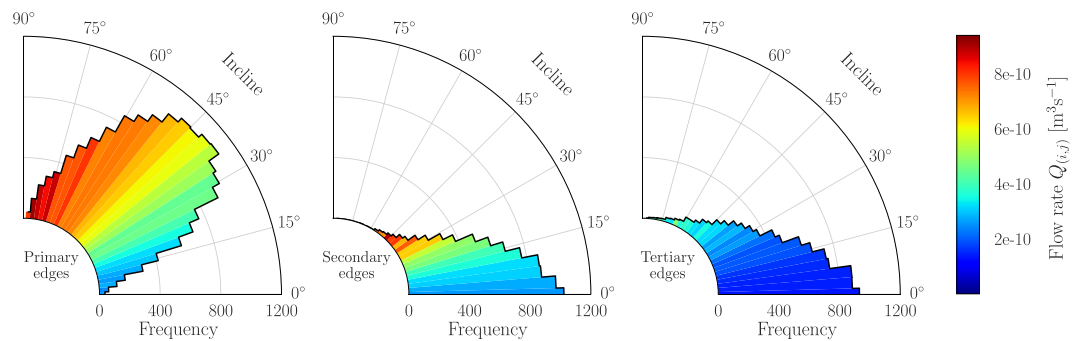
Figure 5. Secondary pathways are poorly aligned with the pressure gradient, yet exhibit favorable flow rates, suggesting an important role in bridging preferential flow pathways. Radial distribution of the incline of primary edges (left), secondary edges (middle) and tertiary edges (right). The color of each bar indicates the average flow rate of all edges in the corresponding range of inclines.

flow rates than tertiary edges. Further qualitative examination of these pathways, shown on the right-hand side of Fig. 1, reveals that secondary edges are either *intra*connections or *inter*connections for the preferential pathways of primary edges. Intraconnections are embedded in chains of primary edges, forming preferential flow pathways that span the assembly from inlets to outlets and carry the fluid across. Interconnections, in contrast, *bridge* these preferential pathways, providing the ability to re-route flow in order to maximize the overall throughput. Interconnecting secondary edges therefore play a crucial role in the global optimization of finding pathways that maximize flow while minimizing cost.

Impact of pathway hierarchy on permeability. The importance of inter- and intra-connecting flow is further highlighted by removing all secondary edges from the original pore network and re-calculating the permeability shown in Fig. 3. To do so, we repeat the process described in the previous section, fitting a logistic regression classifier for every assembly in the loading history and identifying primary, secondary and tertiary edges. In Fig. 6(a), the impact on permeability by removing secondary edges is compared to the impact on the permeability when tertiary edges are removed instead. Removing primary edges disconnects the inlets from the outlets altogether, hence a permeability for this type of removal is not shown.

We observe a pronounced difference of the impact on permeability if secondary edges are removed, compared to removing tertiary edges. On average, permeability is reduced by approximately 10% throughout the loading history if tertiary edges are removed. Prior to peak stress, the reduction is 30% for secondary edges, increasing to a 37% reduction in the steady state. It should be noted that the number of secondary edges is generally slightly smaller than the number of tertiary edges in each assembly, with an average of 7914 and 8929 edges across all strain states, respectively. The per-edge percentual reduction in permeability is still consistently and considerably larger for secondary edges, however: $-33.6\%/7914 = -4.2 \cdot 10^{-3}\%$ for secondary edges, versus $-9.9\%/8929 = -1.1 \cdot 10^{-3}\%$ for tertiary edges.

Impact of pathway hierarchy on transmission efficiency. We quantify the association between permeability, preferential flow, the primary-secondary-tertiary distinction and network connectivity patterns using measures of complex network centrality. We calculate the average *closeness centrality*, *betweenness centrality* and *information centrality*. Closeness and betweenness are based on shortest paths, assuming that flow occurs only along these shortest paths and does not split. More specifically, an unweighted shortest path between node i and j is the route through the pore network that requires the minimum number of edges. The closeness centrality of a node i, in turn, is the reciprocal of the summed shortest path lengths to every other node in the pore network. A larger average closeness centrality suggests that most nodes in the network are relatively well-connected to all other nodes. Edge betweenness centrality is defined as the fraction of all shortest paths between pairs of nodes in the pore network that pass through the edge under consideration. In this case, only the paths that connect an inlet node with an outlet node are considered. Large average betweenness centrality indicates that shortest paths through the pore network are relatively localized and less spread out. Information centrality models connectivity through electrical current flow paths that can split and may follow any (potentially sub-optimal) route. This type of centrality is calculated for a node i as the reciprocal of the summed potential differences between i and every other node j. The edge resistance, a required input of the algorithm, is conveniently calculated as the inverse of the conductance capacity used in our network flow model. High average information centrality, in our case, suggests a more well-connected network in terms of *all* possible paths between nodes.

Figure 6(b–d) illustrates the impact of removing all secondary or all tertiary edges from the pore network. Betweenness centrality is edge-averaged, while closeness- and information centrality are averaged across all nodes.

The betweenness centrality results in Fig. 6(b) show that shortest paths between inlets and outlets gradually become more localized following peak-stress. More interestingly, shortest paths between inlets and outlets become considerably more localized when secondary edges are removed, compared to removing tertiary edges, up to a factor two at 30% axial strain. The removal of inter-connecting edges would contribute to such localization, eliminating the capacity to re-route and focusing shortest paths down the remaining routes connecting
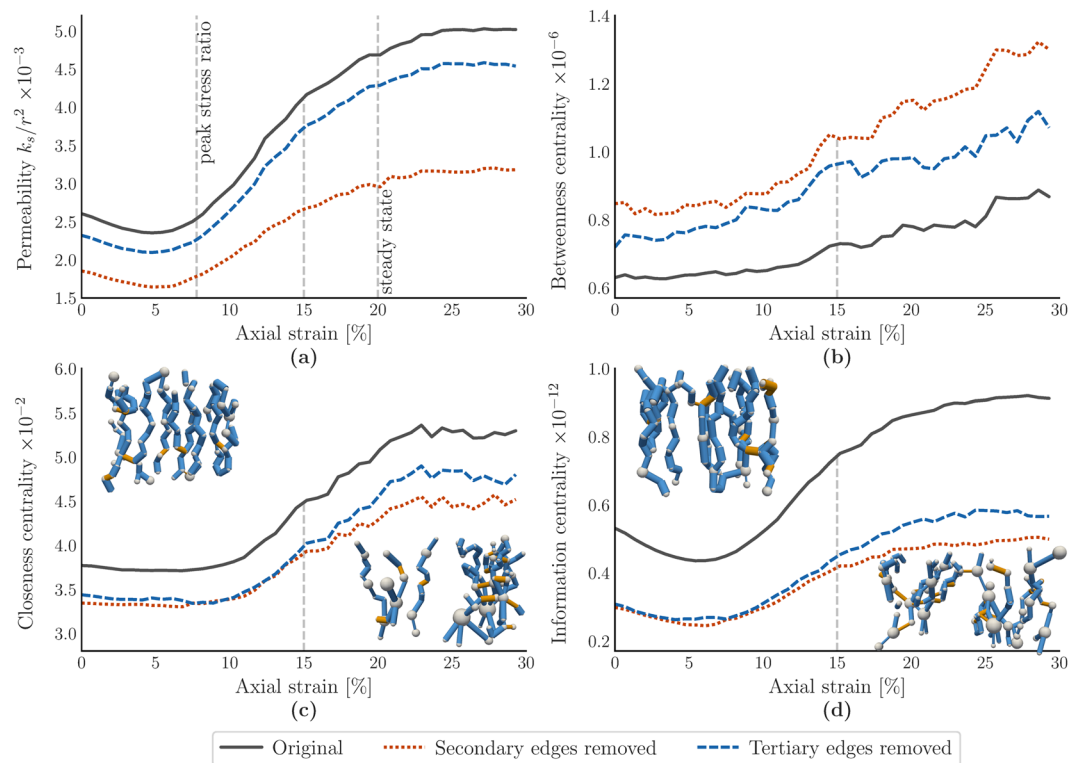
Figure 6. Removing secondary, bridging pathways has a stronger negative effect on permeability and centrality than removing tertiary pathways. Moreover, secondary pathways become more important for the overall transmission beyond peak stress and into the steady state. Impact of removing secondary or tertiary edges on (**a**) permeability, (**b**) edge-averaged betweenness centrality for all shortest paths between inlets and outlets, (**c**) node-averaged closeness centrality for all shortest paths, (**d**) node-averaged information centrality. Insets show pre- and post-peak stress excerpts of the 10% lowest-cost pathways in P, with nodes and edges sized by pore size and throat size, respectively. Primary edges are colored blue and secondary edges are colored orange.

inlets with outlets. The difference in closeness centrality by removing secondary and tertiary edges in Fig. 6(c) is negligible until 15% axial strain. Beyond 15%, removing secondary edges decreases closeness centrality more considerably. The transition suggests that inter- and intra-connections, as provided by secondary edges, become more important in terms of maintaining connectivity as the sample fails. Indeed, the impact on betweenness centrality in Fig. 6(b) of removing secondary edges is also more pronounced in the steady state, compared to prior axial strain values.

Lastly, the trend in information centrality (Fig. 6(d)) is remarkably similar to the trend in permeability. This may not be entirely surprising, given the use of conductance in both results, and the corresponding similarities between calculating pressure differences using Stokes' law and electrical potential differences using Kirchhoff's laws. The Pearson correlation coefficient between the permeability and closeness centrality is equal to 0.975 ($p < 2 \cdot 10^{-16}$), compared to 0.995 ($p < 2 \cdot 10^{-16}$) for the correlation between permeability and information centrality. Note that the decrease in closeness centrality when secondary/tertiary edges are removed is smaller than the decrease in information centrality. We attribute this difference to the fact that information centrality counts 'distance' as an aggregated amount of information flow along *all possible paths*. Closeness centrality only considers the shortest path. As a result, removing edges is more impactful for information centrality. The impact of removing secondary edges in terms of information centrality is more pronounced after 15% axial strain. The decrease in shortest-path connectivity (Fig. 6(c)) and connectivity of all paths (Fig. 6(d)) through removal of secondary edges may contribute to the 37% permeability reduction (Fig. 6(a)) in the steady state, up from 30% prior to peak stress.

Impact of deformation on pathway hierarchy.   Finally, we employ the proposed definition of preferential flow and the resulting hierarchy of primary-secondary-tertiary flow pathways to shed light on the impact of deformation on preferential flow pathway connectivity and structure. First, we investigate the frequency of primary and secondary edges, as well as the structures that they form. To this end, we introduce the notion of *primary components* as the disconnected components in P if all but the primary edges are removed. Similarly, the *secondary* components are the disconnected components in P if only secondary edges are retained. Figure 7 compares the frequency of primary and secondary components with the relative frequency of the primary and secondary edges. Figure 7(a,b) show that a decrease in the percentage of primary edges coincides with a slight increase in the relative frequency of secondary edges. The largest connected component, quantified in size in Fig. 7(d), connects 80–90% of the primary edges at 0% axial strain, in addition to the approximately 1000 other primary components shown Fig. 7(c). The subsequent decrease in size of the largest component coincides with
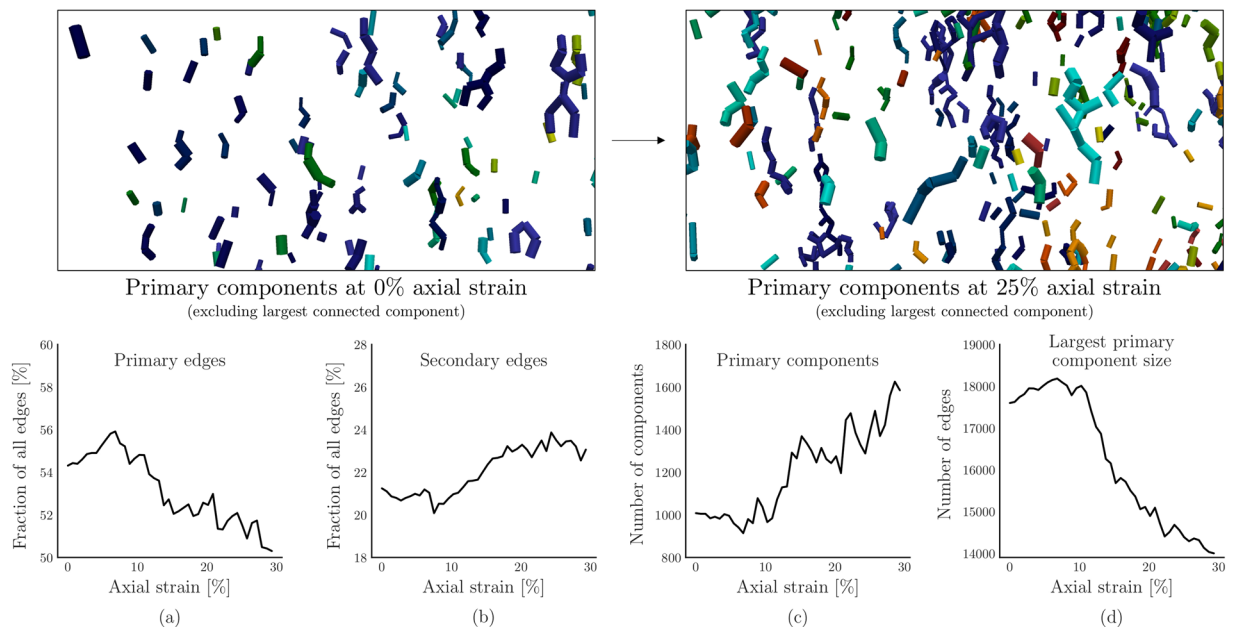
Figure 7. A decrease in the fraction of primary edges coincides with an increase in the number of primary components, indicating an increasingly fragmented primary flow network in the steady state. Showing (**a**) fraction of primary edges, (**b**) fraction of secondary edges, (**c**) number of primary components, and (**d**) number of secondary components. Insets at the top illustrate the primary components at 0% and 25% axial strain.
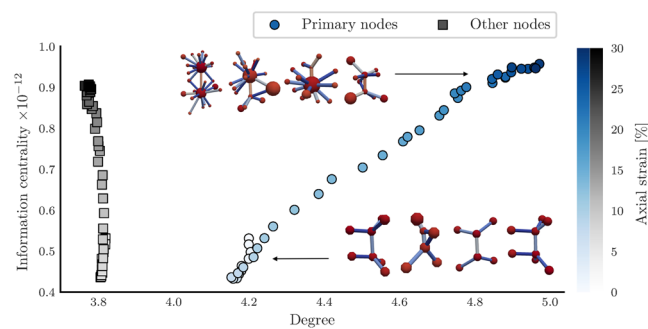

Figure 8. During deformation, and particularly in the steady state, nodes in primary pathways of flow become more connected to other nodes, both locally, in terms of degree, and across multiple scales, in terms of information centrality. Non-primary nodes also gain in multi-scale connectivity, but do not become more locally connected. Feature vector trajectory, in terms of degree and information centrality, for primary and non-primary nodes. Color gradient indicates the axial strain. Insets illustrate nodes with some of the highest centrality values prior to peak stress, and in the steady state.

an increase in the number of primary components. From this, we derive that the primary flow structures become increasingly fragmented as the assembly reaches the steady state. Given the favorable flow characteristics of primary edges, shown in Fig. 5, this fragmentation would have a negative impact on permeability. Yet the permeability increases with increasing axial strain, as shown in Fig. 3(b). We attribute this counter-intuitive behavior to the increasingly prominent role of secondary edges. The frequency of these inter- and intra-connecting edges does increase slightly, as shown in Fig. 7(b). More importantly, the results in Fig. 7 provide further evidence for the idea that secondary edges provide re-routing capabilities for the (increasingly fragmented) primary components, as the sample deforms, corroborating similar observations in the centrality measures presented in Fig. 6.

Figure 8 compares the evolution of local connectivity in terms of degree (i.e. pore coordination number), with changes of multi-scale connectivity in terms of information centrality, as the axial strain increases. Each point represents an average across all nodes in the stated category. A node is considered primary if its primary degree (number of adjacent primary edges) is larger than the global average primary degree across all nodes. Generally, the ratio of primary nodes to other nodes is approximately 1:2.

For primary nodes, local connectivity increases as the axial strain increases, from an average of 4.2 connected edges prior to peak stress to an average degree of 5 in the steady state. The average degree of non-primary nodes, in contrast, remains constant at 3.8 edges, despite continued physical deformation of the material. Both primary and non-primary nodes become more connected to all other nodes in terms of information centrality post peak

stress, with primary nodes achieving slightly higher values towards the end of the shear test. Closer inspection of the associated network structures (insets in Fig. 8) reveals that primary nodes post peak stress, and into the steady state, become highly-connected, micro-structural information (i.e. fluid) hubs. This connectivity may be associated with the above-average flow rates for primary edges (connected to these primary nodes) and below-average flow rates for tertiary edges in Fig. 5. The trend in degree for primary nodes also suggests a correlation with permeability, which is found to be equal to 0.982 ($p < 2 \cdot 10^{-16}$) in terms of the Pearson correlation coefficient, similar to the correlation between permeability and information/closeness centrality.

Discussion

In this section, we connect our insights with existing definitions of preferential flow and prior experimental work. The proposed method to identify preferential flow pathways and corresponding functional groups extends existing interpretations in the following ways. As outlined in the introduction, preferential flow has previously been defined by the delineation of the liquid phase in relatively mobile and immobile parts[1], where the preferential (mobile) part bypasses a fraction of the porous matrix[2], through large inter-aggregate voids also known as macropores[3]. Indeed, in Fig. 3(a), we observe good correlation between what the network flow model predicts to be the fast-flowing, high-throughput fraction of the pore network, and what Stokes flow predicts as such. As a result, the edge flow $f_{(i,j)}$ provided by MFMC may be interpreted as micro-structural quantification of confidence that the edge is preferred, in addition to the macro-structural confidence parameters provided in Table 1). We also observe a statistically significant difference in the average flow rate through primary edges ($Q_{(i,j)} = 5.9 \cdot 10^{-10}$ m^3 s^{-1}) and tertiary edges ($Q_{(i,j)} = 2.1 \cdot 10^{-10}$ m^3 s^{-1}) in Fig. 5. The secondary edges exhibit an intermediate average flow rate of $4.0 \cdot 10^{-10}$ m^3 s^{-1}, however, suggesting an extension of the mobile-immobile distinction to include a third part for the bridging pathways that provide re-routing capabilities for the primary preferential pathways and maximize overall throughput, as hypothesized in the introduction. Secondary interconnections can be observed in the experimentally observed preferential fluid flow pathway images of Ju *et al.*[12], though they are not fully captured by the shortest path model presented therein. Shortest paths alone do not account for pathway capacity and the corresponding inherent interplay between redundancy and rerouting in complex systems. As a result, the secondary (bridge-like) connections between primary preferential flow pathways observed in X-ray CT scans of soil columns[19] cannot be captured using shortest paths alone. Addressing this issue, MFMC takes both cost (shortest paths) and capacity into account, from which the distinction between primary and secondary (bridging) preferential route segments emerges (e.g. see Fig. 1). In capturing primary, secondary and tertiary pathways of preferential fluid flow, our MFMC results corroborate the self-organizing, complex linkage continuum observed by Sidle *et al.*[16] and Luo *et al.*[19].

Our definition of preferential flow also corroborates the well-known preference for larger (macro)pore sizes[3], as indicated by the positive odds (1.36) of the pore volume term in the logistic regression model (Table 1). Rather than capturing preferential flow using individual parameters such as (macro)pore size, our model considers a multitude of multiscale factors influencing preferential flow simultaneously (pore size[3], throat area[12], path alignment with the pressure gradient and surrounding pathways[16,19]) to obtain the optimal decision boundary between primary and secondary pathways. As such, the classifier provides a direct statistical interpretation for other relevant factors as well: the odds of an edge being included in the preferential flow network decrease by a factor 0.55 for every unit increase in edge cost, for example, and increase by a factor 2.59 for every unit increase in edge incline. Interpreted as such, the logistic regression model provides deeper insights into why certain parts of the porous medium are predicted as being part of the preferential flow network by our network flow model.

In the observations of inter- and intra-connecting edges summarized in Fig. 1, a particular analogy emerges with the two-tier hierarchical architecture observed in contact networks and force chains. Prior to peak stress, the sample is compressed along the vertical axis. The material can sustain the applied load through the formation of what is known as the *strong*, load-bearing force network along particle contacts, laterally supported and bridged by contacts in the *weak* force network[38,57]. A loss of lateral support and the phenomenon of *buckling force chains* results in subsequent failure of the assembly post peak-stress[51]. In rock formations, such events are associated with the formation of fracture patterns[53]. Figure 3(b) and 3 show that permeability, maximum flow and total cost are approximately constant in the critical regime, aside from minor fluctuations associated with local compression and dilation resulting from the spatiotemporal formation and buckling of force chains supporting the material. The plateauing behavior of the maximum flow in the steady state has also been observed in force transmission studies using MFMC[36]. In the same way that force is distributed under loading by forming a hierarchical support structure, our observations suggest that preferential fluid flow maximizes throughput by bridging primary flow pathways with secondary, horizontally-oriented connections. As such, juxtaposing our work with the fracture pattern model of Ord and Hobbs[53], we observe that preferential flow is routed not just between macropores *within* regions of dilation (i.e. fracture patterns), but also, through secondary (multiscale) connections, *between* such regions. The supporting structures for force transmission and preferential flow are also inherently connected: compression and dilation due to the formation and buckling of force chains continuously alter the pore space[10,37]. Indeed, Fig. 6 shows that the secondary connections between primary flow pathways become increasingly important for preferential flow as the assembly dilates beyond peak-stress, furthering the idea of Allaire *et al.*[4] that lateral flow connections are at least as important as preferential flow in the direction of gravity (for overall throughput). Additionally, the loss of inter- and intra-connecting edges analyzed in Fig. 6 demonstrates that secondary pathways are responsible for a larger fraction of the connectivity and physical flow than tertiary edges.

Related work has shown that closeness centrality is more strongly correlated with permeability than most other complex network parameters, including betweenness centrality[54,55]. Building on this outcome, the results in Figs. 6 and 8 suggest that information centrality, previously not considered due to computational costs, is (linearly) more strongly correlated with permeability than closeness centrality. We noted in the discussion of Fig. 8 that permeability is also significantly correlated with the average degree of primary nodes, yet not correlated with

the average degree of non-primary nodes. While we cannot prove causality, these results suggest that the local increase in connectivity of primary nodes (depicted in the insets of Fig. 8) may be responsible for the increase in multi-scale connectivity (closeness centrality and information centrality) and permeability post peak stress. Translating these insights back to the physical domain, increased local connectivity of primary nodes implies that pores likely to be part of preferential flow pathways become (on average) more connected to neighboring pores during deformation of the material. The degree of pores outside the preferential flowpaths, in contrast, does not increase. Further inspection shows that primary nodes are likely to represent larger pores, as predicted by the logistic regression model ($e^{\beta_3} = 1.36$ in Table 1) and consistent with the macropore definition of preferential flow[3]. This insight is also consistent with observations by Ju et al.[12], who observed a preferential flow path preference for macropores both in experimental measurements and their shortest path model. Large pores, in turn, are the result of local dilation and shear bands that follow the aforementioned buckling force chains post peak-stress[37,51,53]. Additionally, we observe in Fig. 6 that the negative impact on permeability and connectivity by removing secondary edges (bridging primary nodes/edges) is most pronounced in the steady state. We conclude that preferential flow benefits from the onset of failure at peak-stress, a result that was previously hypothesized in 2D by Russel et al.[10]. While primary, flow-promoting edges decrease in number and become increasingly fragmented post peak-stress, Fig. 7 shows that secondary edges compensate for this by providing the inter- and intra-connecting routes that maximize the throughput, ultimately resulting in an increase in permeability.

This work is a first step in defining and predicting preferential flow structure for more complex materials using MFMC. For saturated media, as considered in this work, comparing MFMC to Navier-Stokes fluid flow[56] and considering coupled interactions (e.g. pore pressure, internal erosion and fault slip[15]) could provide further insights into the applicability and limitations of MFMC, and the structure of preferential fluid flow pathways. Particularly in the presence of higher Reynolds' numbers and stronger inertia effects, the contribution and role of secondary edges may become more or less pronounced. Beyond the assumption of full saturation, hydrological effects in unsaturated media, typically modelled by Richards equation[23], should also be investigated, including wettability effects, drainage and air entrapment[19]. In our idealized particle assemblies, the network flow model highlights non-linear interdependencies between microstructure and preferential path formation in fully saturated granular media. For real granular materials, such non-linear and non-intuitive responses will likely be even more pronounced. A comparison to real fluid flow obtained from MRI scans or X-ray tomography[50], similar to the work by Ju et al.[12] for shortest paths, could provide experimental (network and capacity) input and validation for MFMC. The MFMC framework is exemplified herein with idealized realistic virtual samples and its resulting insights are contrasted with relevant observations available in literature. Beyond structure, applying MFMC to real pore geometries may also provide further insights into the physical dimensions of preferential flow pathways. As long as a pore network representation with corresponding capacities is available, MFMC is generally applicable to void descriptions in permeable media.

Methods

Discrete element simulation.
The discrete element model used here is from Pucilowski et al.[48,49]. It is a three-dimensional analogue of a well-studied in-plane biaxial compression test[51] under a relatively high constant confining pressure to induce dilation during shear. Indeed, these samples were designed for characterization of dynamics of transmission in a densely-packed 3D granular sample up to and during failure when the sample deforms in the presence of fully developed dilatant shear zones. As shown by Ord[52], fluid flow is enabled by regions of dilatancy (e.g. shear zones). The model parameters are summarized in Table 2. The height coordinate is z and the width and depth coordinates are x and y, respectively. In summary, a normally consolidated drained triaxial test is simulated for a dry, sand-like specimen. As is common in discrete element methods, a friction-free approach is used to simulate air pluviation. 6,776 spherical particles are placed in a friction-free, stress-free, non-contacting, and gas-like state in a cubic box without gravity. The walls are subsequently moved inwards to compress the sample to 495 kPa constant confining pressure. When the assembly reaches a static equilibrium, particle velocities are reset and particles are assigned the friction coefficients listed in Table 2. The horizontal (top and bottom) walls are assigned the same sliding friction coefficient as the particles (0.7), while vertical walls in the remaining simulation are still assumed to be frictionless. This way, no additional artificial shear is imposed to the virtual specimen. Next, the assembly is subjected to a triaxial shear test, with a constant axial strain rate of -0.05 s^{-1} in the vertical direction. During the test, the assembly expands globally in horizontal directions under constant confining pressure. The state of the dry assembly is extracted at 430 equally spaced iterations, corresponding to 430 stages of axial strain between 0 and 30%, for further analysis.

Pore network construction.
We use the modified Delaunay tessellation (MDT) approach to construct the pore networks, originally proposed by Al-Raoush et al.[58], with the adaptations introduced by van der Linden et al.[54]. For a detailed description of the algorithm and a comparison with the original MDT approach, refer to[55]. The Delaunay tessellation delineates the pore space using tetrahedra, for which the vertices coincide with the particle centroids. The tetrahedra are assumed to encapsulate pores, while throats are assumed to coincide with the tetrahedral faces. Al-Raoush et al.[58] recognized that the tessellation may unnecessarily sub-divide pores, however, and introduced the idea of merging tetrahedra to remedy this issue. We merge a pair of tetrahedra if the areal porosity on the shared face exceeds a pre-defined threshold[54]. The threshold is fixed at 0.4, which has been shown to achieve a good balance between under- and over-merging[55]. Following the merging procedure, network nodes are assigned to individual or merged collections of tetrahedra, connected by an edge if the tetrahedra share a face. Nodes are labelled as inlet (outlet) nodes if one or more tetrahedra contains a face exposed to the top inlet (bottom outlet) plane.

Number of particles	6,776
Particle radii (from uniform distribution)	$(0.76 - 1.52) \times 10^{-3}$ m
Density	2650 kg/m^3
Young's modulus	9.19×10^7 Pa
Poisson's ratio	0.5
Rolling stiffness coefficient[37]	1.0
Rolling friction coefficient	0.7
Sliding friction coefficient	0.02
Strain rate	$-0.05\, s^{-1}$
Constant confining pressure (in x and z direction)	4.95×10^5 Pa
Initial void ratio	0.5
Initial dimensions ($x \times y \times z$)	$109.23 \times 109.20 \times 109.26\ 10^{-3}$ m
Final dimensions ($x \times y \times z$)	$133.25 \times 132.09 \times 80.97\ 10^{-3}$ m

Table 2. Simulation parameters used in DEM.

Pore volume V_i for node i is calculated by summing the void volume encapsulated by the corresponding tetrahedron or tetrahedra. Similarly, throat area $A_{(i,j)}$ is obtained by summing the void area on all shared faces. The derived variables of adjacent average pore volume and edge incline in Table 1 are defined as:

$$\overline{V}_{(i,j)} = (V_i + V_j)/2 \tag{1}$$

$$\theta_{(i,j)} = \frac{180}{\pi} \arcsin \frac{y_i - y_j}{|(x_i - x_j, y_i - y_j, z_i - z_j)|}, \tag{2}$$

where a node i is assumed to have the centroid coordinates (x_i, y_i, z_i).

Network flow model construction. Maximum Flow Minimum Cost (MFMC) is used to model fluid flow between two parallel walls (top inlet and bottom outlet) in a deforming granular material. Our intended application is the identification of preferential flow paths and their joint structure. As such, we formulate the *relative* capacity and cost through measures that capture the capacity of pores and throats to transmit fluid at a certain cost. If, instead, our application was to model actual fluid flow, then MFMC and the *actual* capacity and cost would have to satisfy the assumed physical flow behaviour (e.g. Stokes or Navier-Stokes) and be formulated accordingly. Such a representation is beyond the scope of this work.

In this work, MFMC is implemented using the LEMON library by Dezso *et al.*[59] and formulated as follows. We denote $\mathcal{G} = (V, E, U, C, s, t)$ as the flow network. Vertices (nodes) V represent the pores and the set E contains the edges (links) that represent throats, as described in the previous section on pore network construction. Edges $(i, j) \in E$ transmit a certain flow $f_{(i,j)}$, limited by the capacities $U = \{u_{(i,j)} | (i, j) \in E\}$ and incurring costs $C = \{c_{(i,j)} | (i, j) \in E\}$. The capacity $u_{(i,j)}$ is equal to the maximum number of units of flow $f_{(i,j)}$ that (i, j) can transmit, with $c_{(i,j)}$ representing the cost per flow unit. Flow is injected into $\mathcal{G}$ through an artificial source node s, connected (with infinite capacity and zero cost) to every inlet node. In the same way, a sink node t is connected to every outlet node. Optimization in the MFMC algorithm proceeds in two stages:

1. Find the maximum number of units of flow (*maximum flow*) $|f|^*$ that can be transmitted through $\mathcal{G}$, from source s to sink t, using Goldberg-Tarjan's preflow push-relabel algorithm[60], subject to the constraints:
   $f_{(i,j)} \leq u_{(i,j)} \forall (i, j) \in E$ (capacity constraint)
   $\sum_{i \in V} f_{(i,j)} = \sum_{k \in V} f_{(j,k)}, j \neq \{s, t\}$ (conservation of flow)
   $\sum_{i \in V} f_{(i,t)} = \sum_{j \in V} f_{(s,j)}$ (conservation of flow)
2. Minimize the cost of transmitting $|f|^*$ through $\mathcal{G}$ (*minimum cost*) using the primal network simplex algorithm[61,62], subject to the constraints above.

The first stage of MFMC typically identifies the main bottleneck in the network, restricting $|f|^*$ to the summed capacity of the edges in the bottleneck. The second stage then proceeds to optimize the transmission of $|f|^*$ by finding the minimum cost pathways from the inlet to the bottleneck, and from the bottleneck to the outlet. The solution of the MFMC optimization problem provides (1) the maximum flow $|f|^*$, capturing the total flow that can be 'pushed' through $\mathcal{G}$, (2) a percolating sub-network P of $\mathcal{G}$ that transmits $|f|^*$ at minimum cost, and (3) the total cost of transmitting $|f|^*$, defined as $|c|^* = \sum_{(i,j) \in E} f_{(i,j)} c_{(i,j)}$. We plot the average cost of transmitting a unit of max flow ($|c|^*/|f|^*$) in Fig. 3(c) to reveal how the ease with which the material transmits flow evolves as the material deforms. Lastly, we note that the solution of the network flow model depends on the *relative* values of $u_{(i,j)}$ and $c_{(i,j)}$. That is, if ($|f|^*, |c|^*, P$) is the solution using capacities $u_{(i,j)}$ and costs $c_{(i,j)}$, then ($\alpha|f|^*, \alpha|c|^*, P$) is the solution for capacities $\alpha u_{(i,j)}$ and costs $\alpha c_{(i,j)}$ ($\alpha > 0$).

In the remainder of this section, we present the definitions of capacity and cost. As a proxy for the relative transmission capacity of an edge in the network $\mathcal{G}$, representing a pore-throat-pore series, we assume equivalent cylindrical geometry for the pores and throats. Under the corresponding assumptions of tube flow, conductance is calculated using the Hagen-Poiseuille law:

$$C_\beta = \frac{\pi r_\beta^4}{8\eta L_\beta}, \quad \beta \in \{i, j, t_{(i,j)}\},$$

(3)

where i and j denote the pores on either side of the connecting throat $t_{(i,j)}$, C (m³ Pa⁻¹ s⁻¹) is the conductance, r (m) and L (m) are the respective radius and length the equivalent cylinder, and η (Pa s) is the dynamic viscosity of the fluid. For further details on the derivation of L and r for pores and throats we refer to the more extensive discussion in van der Linden et al.[54,55]. The conductance capacity function is given by the harmonic mean of the pore-throat-pore series:

$$u_{(i,j)} = \frac{L_i + L_{t_{(i,j)}} + L_j}{\frac{L_i}{C_i} + \frac{L_{t_{(i,j)}}}{C_{t_{(i,j)}}} + \frac{L_j}{C_j}}.$$

(4)

For cost, we use a penalty value based on inertia. Capturing the intuition of 'high cost' for flow around a sharp bend, an edge (i, j) is given a high cost value if it aligns poorly with other edges connected to nodes i and j. To this end, we define the average angle of edge (i, j) with the N_k other edges (i, k) connecting to node i as:

$$\overline{\theta_i} = \frac{180}{\pi N_k} \sum_{k=1}^{N} \arccos \frac{\mathbf{v}_{ij} \cdot \mathbf{v}_{ik}}{|\mathbf{v}_{ij}||\mathbf{v}_{ik}|}$$

$$\mathbf{v}_{ij} = \left[x_j - x_i, y_j - y_i, z_j - z_i \right],$$

$$\mathbf{v}_{ik} = [x_k - x_i, y_k - y_i, z_k - z_i],$$

(5)

where (x_i, y_i, z_i) are the coordinates of the centroid of the particle corresponding to node i. The average angle of edge (i, j) with every other edge connecting to node j is obtained by switching out i and j in (5). Using this expression, the inertia cost penalty is given by the harmonic mean:

$$c_{(i,j)} = \left[\frac{2}{\overline{\theta_i}^{-1} + \overline{\theta_j}^{-1}} \right]^{-1}$$

(6)

We take the inverse of the harmonic mean to ensure angles close to 180°, in alignment with (i, j), result in low cost.

Classification. To perform the classification, we define a simple binary classifier, predicting inclusion or exclusion $\chi_{(i,j)}$ of a particular edge in the MFMC sub-graph P, i.e. $\chi_{(i,j)} = 1$ if $(i, j) \in P$ and $\chi_{(i,j)} = 0$ otherwise. The classifier is given by the logistic regression of $\chi_{(i,j)}$ on the four independent edge-level variables (features) $X_1, \ldots, X_4$ listed in Table 1, estimating the corresponding model parameters $\beta_0, \ldots, \beta_4$ by maximum likelihood in:

$$logit(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4, \quad logit(p) = \ln\left(\frac{p}{1-p}\right),$$

(7)

where p is the probability that $\chi_{(i,j)} = 1$. Other features were considered (capacity, throat length, vertical distance) but ultimately removed from the model due to multicollinearity given by other features in Table 1. Other (non-linear) classification methods were also considered, but did not provide noticeable improvement in classification performance. The features $X_1, \ldots, X_4$ are standardized to have zero-mean and unit-variance.

For previously unseen edges, a derived output of the logistic regression model in Eq. (7) is the probability that $\chi_{(i,j)} = 1$. As such, a threshold is required to convert each probability prediction into the binary outcome: $(i, j) \in P$ or $(i, j) \notin P$. For probabilities above the threshold, we assume $\chi_{(i,j)} = 1$ (positive case), while probabilities below the threshold are assumed to imply $\chi_{(i,j)} = 0$ (negatives case). As is common, we calculate this threshold by simultaneously maximizing the fraction of true positives (*sensitivity*) and fraction of true negatives (*specificity*) using 20% of the 80% of edges reserved for model fitting in the sample with 21.54% axial strain used for Table 1. Sensitivity and specificity are competing objectives, as increasing the threshold increases the likelihood of correctly predicting all negative cases, while decreasing the likelihood of predicting all positive cases. For this particular set of edges, we obtain a threshold of 0.73.

Permeability. The approximated numerical permeability is calculated using Stokes' law and the OpenPNM Python package[63]. Using the conductances $C_{(i,j)}$ calculated in Eq. (4), the conservation of mass prescribes:

$$\sum_{i \to j} Q_{(i,j)} = \sum_{i \to j} C_{(i,j)} (P_i - P_j) = 0,$$

(8)

where $Q_{(i,j)}$ is the flowrate (m³ s⁻¹) between nodes i and j and P_i is the pressure (Pa). We express Eq. (8) as linear system of equations, prescribing $P_i = 10$ Pa for all inlet nodes i, and $P_j = 9$ Pa for all outlet nodes j. We obtain the pressures at all other nodes by solving the corresponding matrix-vector equation. Lastly, the net outlet flow Q_{our} is inserted into Darcy's law to calculate permeability:

$$k = \frac{\eta L Q_{out}}{A \Delta P},$$

(9)

where L is the height of the sample, A is the cross-sectional area, $\eta = 0.001002$ Pa s is the fluid dynamic viscosity, and $\Delta P = 1$ is the pressure difference over inlet and outlet nodes. For spherical particle assemblies, Stokes flow has been shown to agree moderately well with physical measurements[64].

Shortest paths and network centrality. We define $d_{(i,j)}$ as the length of the shortest path distance from node i to node j, minimizing the number of traversed edges. Shortest paths are calculated using Dijkstra's algorithm[65]. To assess network connectivity, we compute the average closeness, betweenness and information centrality. Closeness centrality for a node i is given by[66]:

$$C_i^C = (|V| - 1)\left[\sum_{j \neq i} d(i, j)\right]^{-1},$$

(10)

where $|V|$ is the number of nodes. Closeness centrality is a multi-scale measure of connectivity, capturing the degree of connectivity of nodes to all other nodes. Betweenness centrality further characterizes this connectivity by computing the fraction of shortest paths passing through a particular edge[67]:

$$C_{(i,j)}^B = \frac{1}{2}(|V| - 1)(|V| - 2) \sum_{k \in V_{in},\ l \in V_{out}} \frac{\sigma_{kl}(i, j)}{\sigma_{kl}},$$

(11)

where V_{in} and V_{out} are the respective inlet and outlet nodes, σ_{kl} is equal to the total number of shortest paths between node k and l, and $\sigma_{kl}(i, j)$ is the number of shortest paths between k and l that pass through edge (i, j). We opt for the specific case where $k \in V_{in}$ and $l \in V_{out}$ (as opposed to $k, l \in V$) to assess the degree of path localization in the direction of flow.

Shortest path distance, closeness centrality and betweenness centrality quantify connectivity by assuming quantities traverse the network solely along non-diverging shortest paths. In a connected and saturated porous medium, however, fluid will flow along non-preferential paths too. To this end, information centrality quantifies connectivity through electrical current flow paths, which, similar to electrical current, may diverge, converge and follow non-preferential paths[68]. We calculate information centrality for every node i as[69]:

$$C_i^I = (|V| - 1)\left[\sum_{j \neq i} p_{ij}(i) - p_{ij}(j)\right]^{-1},$$

(12)

where $p_{ij}(k)$ is the absolute potential of node k if a single unit of current is injected at node i and leaves the network at node j. Given this unit supply, the absolute potential is calculated by solving the linear set of equations given by Ohm's and Kirchoff's potential laws:

$$\sum_{i \to j} x_{(i,j)} = \sum_{i \to j} C_{(i,j)}(p_i - p_j) = 0,$$

(13)

where $x_{(i,j)}$ is the current and $C_{(i,j)}$ is the conductance. We use (4) for the conductance, in which case (13) is equivalent to Stokes law' (8) with a unit flow supply at node i and sink at node j. The term $p_{ij}(i) - p_{ij}(j)$ in (12) is the effective resistance (or pressure difference) acting as an alternative to the shortest path distance used in (10).

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Simunek, J., Jarvis, N. J., van Genuchten, M. T. & Gardenas, A. Review and comparison of models for describing non-equilibrium and preferential flow and transport in the vadose zone. *J. Hydrol.* **272**, 14–35 (2003).
2. Hendrickx, J. & Flury, M. *Uniform and Preferential Flow Mechanisms in the Vadose Zone* (The National Academies Press, Washington, DC, 2001).
3. Jarvis, N. J. A review of non-equilibrium water flow and solute transport in soil macropores: Principles, controlling factors and consequences for water quality. *Eur. J. Soil Sci.* **58**, 523–546 (2007).
4. Allaire, S. E., Roulier, S. & Cessna, A. J. Quantifying preferential flow in soils: A review of different techniques. *J. Hydrol.* **378**, 179–204 (2009).
5. Sammartino, S. *et al.* Identifying the functional macropore network related to preferential flow in structured soils. *Vadose Zone J.* **14**, 1–16 (2015).
6. Koestel, J. K., Moeys, J. & Jarvis, N. J. Evaluation of nonparametric shape measures for solute breakthrough curves. *Vadose Zone J.* **10**, 1261–1275 (2011).

7. de Vries, E. T., Raoof, A. & van Genuchten, M. T. Multiscale modelling of dual-porosity porous media; a computational pore-scale study for flow and solute transport. *Adv. Water Resour.* **105**, 82–95 (2017).

8. Miao, X., Narsilio, G., Wu, A. & Yang, B. A 3d dual pore-system leaching model. part 1: Study on fluid flow. *Hydrometall.* **167**, 173–182 (2017).

9. Yang, G., Cook, N. & Myer, L. Analysis of preferential flow paths using graph theory. *Int. J. Rock Mech. Min. Sci. & Geomech. Abstr.* **30**, 1423–1429 (1993).

10. Russell, S., Walker, D. & Tordesillas, A. A characterization of the coupled evolution of grain fabric and pore space using complex networks: Pore connectivity and optimized flows in the presence of shear bands. *J. Mech. Phys. Solids* **88**, 227–251 (2016).

11. Rizzo, C. B. & de Barros, F. P. Minimum hydraulic resistance and least resistance path in heterogeneous porous media. *Water Resour. Res.* **53**, 8596–8613 (2017).

12. Ju, Y. *et al.* Prediction of preferential fluid flow in porous structures based on topological network models: Algorithm and experimental validation. *Sci. China Technol. Sci.* 1–11 (2018).

13. Jarvis, N., Koestel, J. & Larsbo, M. Understanding preferential flow in the vadose zone: Recent advances and future prospects. *Vadose Zone J.* **15**, 0 (2016).

14. Gerke, H. H. Preferential flow descriptions for structured soils. *J. Plant Nutr. Soil Sci.* **169**, 382–400 (2006).

15. Yang, Z. & Juanes, R. Two sides of a fault: Grain-scale analysis of pore pressure control on fault slip. *Phys. Rev. E* **97**, 022906 (2018).

16. Sidle, R. C., Noguchi, S., Tsuboyama, Y. & Laursen, K. A conceptual model of preferential flow systems in forested hillslopes: Evidence of self-organization. *Hydrol. Process.* **15**, 1675–1692 (2001).

17. Holtzman, R. Effects of pore-scale disorder on fluid displacement in partially-wettable porous media. *Sci. Reports* **6**, 36221 (2016).

18. Moreno, L., Tsang, C., Tsang, Y. & Neretnieks, I. Some anomalous features of flow and solute transport arising from fracture aperture variability. *Water Resour. Res.* **26**, 2377–2391 (1990).

19. Luo, L., Lin, H. & Halleck, P. Quantifying soil structure and preferential flow in intact soil using x-ray computed tomography. *Soil Sci. Soc. Am. J.* **72**, 1058 (2008).

20. Lam, M. A. *et al.* The ultrastructure of spinal cord perivascular spaces: Implications for the circulation of cerebrospinal fluid. *Sci. Reports* **7**, 12924 (2017).

21. Rosqvist, H. N., Dollar, H. L. & Fourie, B. A. Preferential flow in municipal solid waste and implications for long-term leachate quality: Valuation of laboratory-scale experiments. *Waste Manag. Res.* **23**, 367–380 (2005).

22. Rosas-Carbajal, M., Komorowski, J. C., Nicollin, F. & Gibert, D. Volcano electrical tomography unveils edifice collapse hazard linked to hydrothermal system structure and dynamics. *Sci. Reports* **6**, 29899 (2016).

23. Weiler, M. Macropores and preferential flow—a love-hate relationship. *Hydrol. Process.* **31**, 15–19 (2017).

24. Le Borgne, T., Bour, O., Paillet, F. L. & Caudal, J. P. Assessment of preferential flow path connectivity and hydraulic properties at single-borehole and cross-borehole scales in a fractured aquifer. *J. Hydrol.* **328**, 347–359 (2006).

25. Matthäi, S. K. & Belayneh, M. Fluid flow partitioning between fractures and a permeable rock matrix. *Geophys. Res. Lett.* **31**, 1–5 (2004).

26. Fredrich, J. T. & Lindquist, W. B. Statistical characterization of the three-dimensional microgeometry of porous media and correlation with macroscopic transport properties. *Int. J. Rock Mech. Min. Sci.* **34**, 368–370 (1997).

27. Biswal, B., Manwart, C., Hilfer, R., Bakke, S. & Oren, P. Quantitative analysis of experimental and synthetic microstructures for sedimentary rock. *Phys. A* **273**, 452–475 (1999).

28. Hilfer, R. & Manwart, C. Permeability and conductivity for reconstruction models of porous media. *Phys. review E* **64**, 021304 (2001).

29. Manwart, C., Aaltosalmi, U., Koponen, A., Hilfer, R. & Timonen, J. Lattice-boltzmann and finite-difference simulations for the permeability for three-dimensional porous media. *Phys. Rev. E* **66**, 016702 (2002).

30. Okabe, H. & Blunt, M. J. Prediction of permeability for porous media reconstructed using multiple-point statistics. *Phys. Rev. E* **70**, 66135 (2004).

31. Xiong, Q., Baychev, T. G. & Jivkov, A. P. Review of pore network modelling of porous media: Experimental characterisations, network constructions and applications to reactive transport. *J. Contaminant Hydrol.* **192**, 101–117 (2016).

32. Lin, Q. & Tordesillas, A. Constrained optimisation in granular network flows: Games with a loaded dice. *AIP Conf. Proc.* **1542**, 547–550 (2013).

33. Tordesillas, A., Cramer, A. & Walker, D. M. Minimum cut and shear bands. *AIP Conf. Proc.* **1542**, 507–510 (2013).

34. Lin, Q. & Tordesillas, A. Towards an optimization theory for deforming dense granular materials. *J. Ind. Manag. Optim.* **10**, 337–362 (2014).

35. Tordesillas, A., Tobin, S. T., Cil, M., Alshibli, K. & Behringer, R. P. Network flow model of force transmission in unbounded and bonded granular media. *Phys. Rev. E* **91**, 1–14 (2015).

36. Tordesillas, A. *et al.* Shear bands as bottlenecks in force transmission. *EPL (Europhysics Lett.* **110**, 58005 (2015).

37. Oda, M. & Iwashita, K. *Mechanics of Granular Materials: An Introduction* (A.A. Balkema, Rotterdam, The Netherlands, 1999).

38. Majmudar, T. S. & Behringer, R. P. Contact force measurements and stress-induced anisotropy in granular materials. *Nat.* **435**, 1079–1082 (2005).

39. Tordesillas, A., Kahagalage, S., Ras, C., Nitka, M. & Tejchman, J. Interdependent evolution of robustness, force transmission and damage in a heterogeneous quasi-brittle granular material: from suppressed to cascading failure. ArXiv Pre-print, 1809.01491 (2018).

40. Ushizima, D. *et al.* Augmented topological descriptors of pore networks for material science. *IEEE Transactions on Vis. Comput. Graph.* **18**, 2041–2050 (2012).

41. Gao, Z.-K., Du, M., Hu, L.-D., Zhou, T.-T. & Jin, N.-D. Visibility graphs from experimental three-phase flow for characterizing dynamic flow behavior. *Int. J. Mod. Phys. C* **23**, 1250069 (2012).

42. Gao, Z. K. *et al.* Multi-frequency complex network from time series for uncovering oil-water flow structure. *Sci. Reports* **5**, 8222 (2015).

43. Gao, Z.-K. *et al.* Pagerank versatility analysis of multilayer modality-based network for exploring the evolution of oil-water slug flow. *Sci. Reports* **7**, 5493 (2017).

44. Newman, M. The structure and function of complex networks. *SIAM review* **45**, 167–256 (2003).

45. Chung, F. & Lu, L. *Complex Graphs and Networks*, vol. 107 (American Mathematical Society, Boston, MA, USA, 2006).

46. Lawyer, G. Understanding the influence of all nodes in a network. *Sci. Reports* **5**, 1–9 (2015).

47. Phillips, J. D., Schwanghart, W. & Heckmann, T. Graph theory in the geosciences. *Earth-Science Rev.* **143**, 147–160 (2015).

48. Pucilowski, S., Tordesillas, A. & Froyland, G. Self-organization in the localised failure regime: metastable attractors and their implications on force chain functionality. *EPJ Web Conf.* **140**, 10007 (2017).

49. Pucilowski, S. & Tordesillas, A. Rattler wedging and force chain buckling: metastable attractor dynamics of local grain rearrangements underlie globally bistable shear banding regime. *Granul. Matter* (2019, to appear).

50. Koestel, J. & Larsbo, M. Imaging and quantification of preferential solute transport in soil macropores. *Water Resour. Res.* **50**, 4357–4378 (2014).

51. Tordesillas, A. Force chain buckling, unjamming transitions and shear banding in dense granular assemblies. *Philos. Mag.* **87**, 4987–5016 (2007).

52. Ord, A. Mechanical controls on dilatant shear zones. In Knipe, R. J. & Rutter, E. H. (eds.) *Deformation Mechanisms, Rheology and Tectonics*, **54**, 183–192 (Geological Society Special Publication, London, 1990).
53. Ord, A. & Hobbs, B. E. Fracture pattern formation in frictional, cohesive, granular material. *Philos. Transactions The Royal Soc. A Math. Phys. Eng. Sci.* **368**, 95–118 (2010).
54. van der Linden, J. H., Narsilio, G. A. & Tordesillas, A. Machine learning framework for analysis of transport through complex networks in porous, granular media: A focus on permeability. *Phys. Rev. E* **94**, 022904 (2016).
55. van der Linden, J. H., Sufian, A., Narsilio, G. A., Russell, A. R. & Tordesillas, A. A computational geometry approach to pore network construction for granular packings. *Comput. Geosci.* **112**, 133–143 (2018).
56. Narsilio, G., Buzzi, O., Fityus, S., Yun, T. & Smith, D. Upscaling of navier–stokes equations in porous media: Theoretical, numerical and experimental approach. *Comput. Geotech.* **36**, 1200–1206 (2009).
57. Radjai, F., Wolf, D. E., Jean, M. & Moreau, J.-J. Bimodal character of stress transmission in granular packings. *Phys. Rev. Lett.* **80**, 61–64 (1998).
58. Al-Raoush, R., Thompson, K. & Willson, C. S. Comparison of network generation techniques for unconsolidated porous media. *Soil Sci. Soc. Am. J.* **67**, 1687–1700 (2003).
59. Dezso, B., Jüttner, A. & Kovács, P. LEMON - an open source C++ graph template library. *Electron. Notes Theor. Comput. Sci.* **264**, 23–45 (2011).
60. Goldberg, A. V. & Tarjan, R. E. A new approach to the maximum-flow problem. *J. ACM* **35**, 921–940 (1988).
61. Dantzig, G. B. *Linear Programming and Extensions* (Princeton University Press, Princeton, NJ, USA, 1963).
62. Ahuja, R. K., Magnanti, T. L. & Orlin, J. B. *Network Flows: Theory, Algorithms, and Applications* (Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993).
63. Gostick, J. *et al.* Openpnm: A pore network modeling package. *Comput. Sci. Eng.* **18**, 60–74 (2016).
64. Kress, J., Yun, T., Narsilio, G., Matthew Evans, T. & Lee, D. Evaluation of hydraulic conductivity in 3d random and heterogeneous particulate materials using network model. *Comput. Geotech.* **40**, 45–52 (2012).
65. Dijkstra, E. W. A note on two problems in connexion with graphs. *Numer. Math.* **1**, 269–271 (1959).
66. Freeman, L. C. A set of measures of centrality based on betweenness. *Sociom.* **40**, 35–41 (1977).
67. Freeman, L. C. Centrality in social networks conceptual clarification. *Soc. Networks* **1**, 215–239 (1978).
68. Stephenson, K. & Zelen, M. Rethinking centrality: Methods and examples. *Soc. Networks* **11**, 1–37 (1989).
69. Brandes, D. & Fleischer, U. Centrality measures based on current flow. In Diekert, V. & Durand, B. (eds.) *Proc. 22nd Symp. Theoretical Aspects of Computer Science* (STACS '05), vol. LNCS **3404**, 533–544 (Springer, Berlin, Heidelberg, 2005).

Acknowledgements

Author contributions

A.T. conceived the main ideas. J.H.v.d.L. developed the code, performed the modeling and wrote the first draft. J.H.v.d.L., A.T. and G.A.N. further refined the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-54699-6.

Correspondence and requests for materials should be addressed to A.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author/s:
van der Linden, JH; Tordesillas, A; Narsilio, GA

Title:
Preferential flow pathways in a deforming granular material: self-organization into functional groups for optimized global transport.

Date:
2019-12-03

Citation:
van der Linden, J. H., Tordesillas, A. & Narsilio, G. A. (2019). Preferential flow pathways in a deforming granular material: self-organization into functional groups for optimized global transport.. Scientific Reports, 9 (1), https://doi.org/10.1038/s41598-019-54699-6.

Persistent Link:
http://hdl.handle.net/11343/233459

File Description:
Published version
License:
CC BY