



**HAL**  
open science

## Statistical metrics for the characterization of karst network geometry and topology

Pauline P. Collon, David Bernasconi, Cécile Vuilleumier, Philippe Renard

► **To cite this version:**

Pauline P. Collon, David Bernasconi, Cécile Vuilleumier, Philippe Renard. Statistical metrics for the characterization of karst network geometry and topology. *Geomorphology*, 2017, 283, pp.122-142. 10.1016/j.geomorph.2017.01.034 . hal-01468055v1

**HAL Id: hal-01468055**

**<https://hal.univ-lorraine.fr/hal-01468055v1>**

Submitted on 15 Feb 2017 (v1), last revised 25 Aug 2021 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Statistical metrics for the characterization of karst network geometry and topology

Pauline Collon<sup>1</sup>, David Bernasconi<sup>2</sup>, Cécile Vuilleumier<sup>1,2</sup>, and Philippe Renard<sup>2</sup>

<sup>1</sup>GeoRessources (Université de Lorraine / CNRS / CREGU), ENSG, 2 rue du doyen Marcel Roubault, BP 10162, 54505 Vandœuvre-lès-Nancy Cedex, France

<sup>2</sup>Centre d'Hydrogéologie et de Géothermie, Université de Neuchâtel, 11 rue Emile-Argand, 2000 Neuchâtel, Switzerland

**Abstract** Statistical metrics can be used to analyse the morphology of natural or simulated karst systems; they allow describing, comparing, and quantifying their geometry and topology. In this paper, we present and discuss a set of such metrics. We study their properties and their usefulness based on a set of more than 30 karstic networks mapped by speleologists. The data set includes some of the largest explored cave systems in the world and represents a broad range of geological and speleogenetic conditions allowing us to test the proposed metrics, their variability, and their usefulness for the discrimination of different morphologies.

All the proposed metrics require that the topographical survey of the caves are first converted to graphs consisting of vertices and edges. This data preprocessing includes several quality check operations and some corrections to ensure that the karst is represented as accurately as possible. The statistical parameters relating to the geometry of the system are then directly computed on the graphs, while the topological parameters are computed on a reduced version of the network focusing only on its structure.

Among the tested metrics, we include some that were previously proposed such as tortuosity or the Howard's coefficients. We also investigate the possibility to use new metrics derived from graph theory. In total, 21 metrics are introduced, discussed in detail, and compared on the basis of our data set. This work shows that orientation analysis and, in particular, the entropy of the orientation data can help to detect the existence of inception features. The statistics on branch length are useful to describe the extension of the conduits within the network. Rather surprisingly, the tortuosity does not vary very significantly. It could be heavily influenced by the survey methodology. The degree of interconnectivity of the network, related to the presence of maze patterns, can be measured using different metrics such as the Howard's parameters, global cyclic coefficient, or the average vertex degree. The average vertex degree of the reduced graph proved to be the most useful as it is simple to compute, it discriminates properly the interconnected systems (mazes) from the acyclic ones (tree-like structures), and it permits us to classify the acyclic systems as a function of the total number of branches. This topological information is completed by three parameters, allowing us to refine the description. The correlation of vertex degree is rather simple to obtain. It is systematically positive on all studied data sets indicating a predominance of assortative networks among karst systems. The average shortest path length is related to the transport efficiency. It is shown to be mainly correlated to the size of the network. Finally, central point dominance allows us to identify the presence of a centralized organization.

## Keywords

karst characterization  
topology  
geometry  
graph theory  
database  
karst pattern

## INTRODUCTION

Several studies have shown the importance of karst network geometry for understanding flow and transport in karstic aquifers [Jeannin, 2001, Chen and Goldscheider, 2014]. But in many cases, the network geometry remains partially, or even totally, unknown. Various simulation methods have thus been developed to tackle this problem. They can be conditioned to field observations and to a partial knowledge of the network that drive orientations and/or length of the conduits, but they hardly consider the global network architecture [e.g., Jaquet et al., 2004, Labourdette et al., 2007, Borghi et al., 2012, Collon-Drouaillet et al., 2012, Pardo-Iguzquiza et al., 2012, Viseur et al., 2014, Hendrick and Renard, 2016]. Metrics that would characterize karstic networks are thus required (i) to

compare artificial and natural networks, and possibly, (ii) to better parameterize the simulation methods.

First attempts to quantitatively characterize karst networks date back to the 1960s [e.g., Curl, 1966, Howard, 1971, Williams and Williams, 1972] and was strongly supported by parallel investigations on river networks [e.g., Horton, 1945, Scheidegger, 1966, 1967, Woldenberg, 1966, Smart, 1969, Howard et al., 1970]. Systematic analysis was then mainly done on two-dimensional planar maps or vertical cross sections produced by speleological explorations [Howard, 1971].

But over the last decade, new exploration and survey tools have emerged that ease the acquisition, storage, and share of three-dimensional data. LIDAR techniques and aerial photography now allow us to rapidly map surface evidence of karst presence [Weishampel et al., 2011, Alexander et al., 2013, Zhu

et al., 2014]. Lidar technology is also used underground and allows now 3D morphological analysis of small portions of conduits and drains [Ployon et al., 2011, Jaillet et al., 2011, Sadier, 2013]. Underground-GPS systems progressively develop [Caverne, 2011], and one can hope that they will therefore facilitate the acquisition of 3D regular data sets on karstic systems, as well as the optical laser device tool recently tested in Yucatan (Mexico: Schiller and Renard [2016]). Combined with the increasing power of computers, these recent advances create a renewal on statistical analysis of karsts, considering now the 3D nature of these systems [Pardo-Iguzquiza et al., 2011, Piccini, 2011, Fournillon et al., 2012]. But in general, these studies focus on the geometrical characterization of the networks, and they rarely compute metrics on more than one or two examples.

In parallel, topological analysis of networks has had an explosive growth, and many new metrics have been proposed that have not yet been applied on natural networks such as karstic systems [Ravasz and Barabasi, 2002, Boccaletti et al., 2006, Costa et al., 2007].

The goals of this paper are (i) to propose a set of metrics to characterize both the geometry and the topology of karstic networks and (ii) to provide a data set of the corresponding values computed on a large ensemble of karstic systems.

With these objectives, we introduce several metrics from graph theory that have not yet been used, to our knowledge, in this context. We compare them with other metrics introduced by previous authors. For each metric, we provide, in addition to its formal definition, some examples on simple networks to help in getting an intuitive understanding of their meaning. We then compute all the metrics on a data set of 34 cave systems gathered thanks to the help of speleologists: 31 networks are real 3D data sets, while 3 come from 2D projection maps of Palmer [1991]. The analysis of these results permit us to compare the metrics, analyse potential correlations, and discuss their relevance for the quantification of karst geomorphology.

## FROM REAL NETWORKS TO GRAPHS

### Data acquisition

A solid statistical analysis would require a data set as large as possible of precise, complete, regular, and homogeneous measurements. Cave mapping is technically difficult and performed thanks to long-term work of trained speleologists. To our knowledge, no centralized database inventories all explored caves and gives open-access to the primary 3D data. To realize this study, various speleologists have been independently contacted and have agreed to share their data. Thus, we collected 31 three-dimensional karst networks from various locations in the world (Table 1). Three 2D networks were also used in the study and complete the database: Blue Spring, Crevice, and Crossroads (USA). Note that four parts of the Sieben Hengste karst (Switzerland) were provided and studied. Subparts *SP1* and *SP2* are included in the *LargePart* network. The *UpPart* is an independent one, located upstream from the *LargePart* network. We have not split nor merged any network parts as we had no accurate information indicating if it should be done and how.

The total explored length of these networks strongly varies: from small networks like Pic du Jer in France, with a total length of around 612 m, to very large ones like Ox Bel Ha in Mexico, with a total explored length of 143 km. The variety of

morphologies of the gathered networks are also interesting to notice. Some are principally developed around some horizons providing them a close to 2D architecture (e.g., Agen Allwed, Foussoubie Goule, Ox Bel Ha), others are characterized by their vertical elongation (e.g., Krubera, Ratasse), and some have developed equally in the three dimensions of space (e.g., Mammuthöhle, Sakany, Sieben Hengste SP1: Fig. 1).

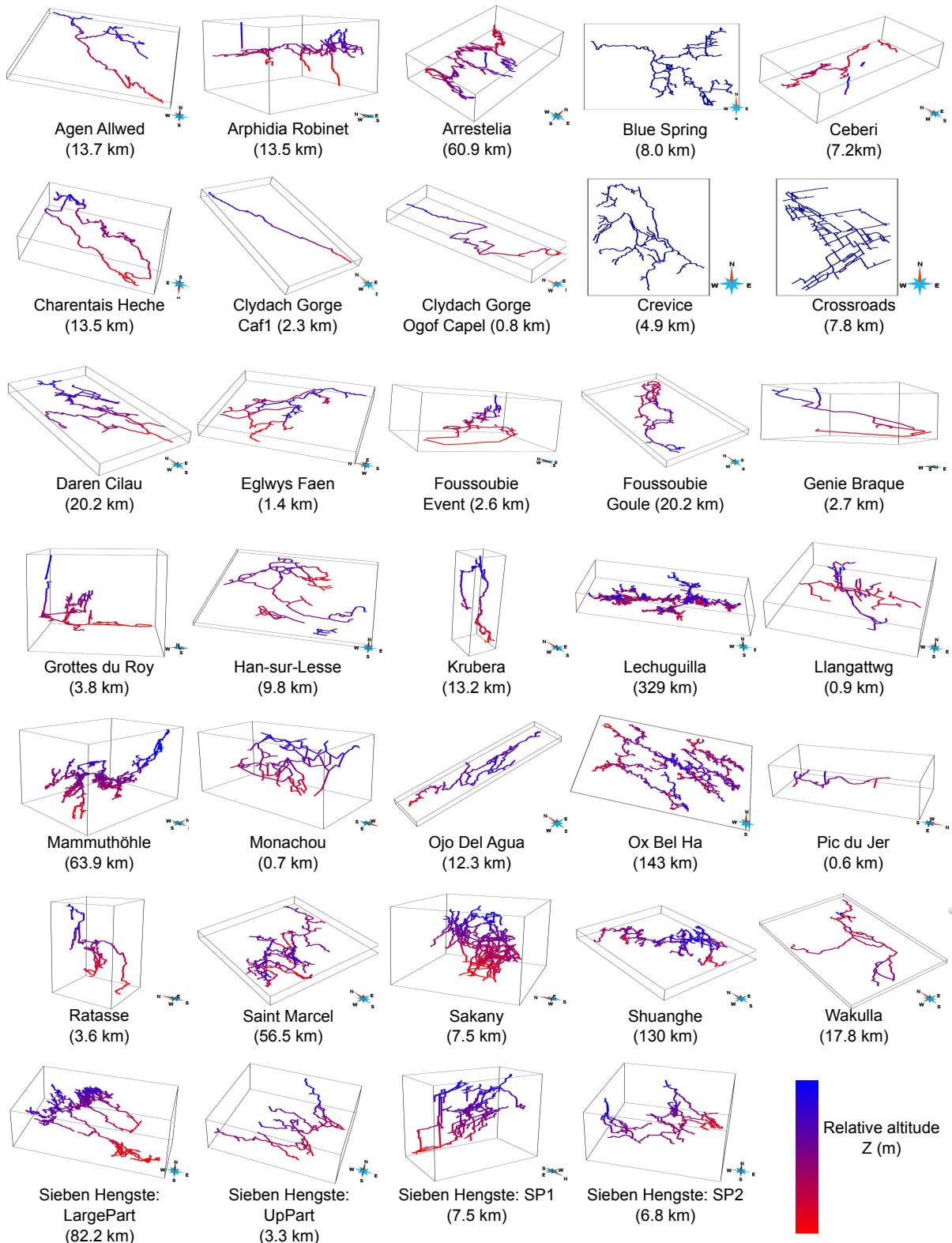
The three 2D networks come from an automated digitalization of 2D maps published by Palmer [1991]. Despite that the data format changes from one source to another, all 3D karst networks mapped by speleologists are available as a sequence of  $n$  topographic stations  $i = 1, 2, \dots, n$  referring to a given origin point  $i = 1$  (e.g., an entrance of the considered cave: Fig. 2A). In general, cave survey data consist in such series of uniquely defined stations linked to each other by lines-of-sight [e.g., Jeannin et al., 2007, Pardo-Iguzquiza et al., 2011]. Surveying methodology varies, but, in general, one line-of-sight,  $c_{ij}$ , linking two consecutive stations  $i$  and  $j$ , is defined by (i) a distance measured with a low-stretch tape or laser range-finder, (ii) a direction (azimuth or bearing) taken with a compass, and (iii) an inclination from horizontal taken with a clinometer (Fig. 2B). Most of the time, but not always, the maximum height and the maximal width of the passage are also measured at each station. Sometimes, it is a more detailed distance to the surrounding walls that is provided through left, right, up, and down measurements [e.g., Jaillet et al., 2011, Rongier et al., 2014].

Despite the recent efforts of the speleological community to homogenise their cave survey methodology and take increasing care of the precision and validity of the field measurements, some survey errors can still appear and are more or less easily detectable:

- Back-sight measures: a station can be measured twice when the surveyor returns in the opposite direction but continues its data acquisition. This can be easily detected when the station is strictly identical, but if the measurement is made just nearby, it can be interpreted as a cycle – also called passage loop (Fig. 3 - case A).
- Cycle closure errors appear when a gap, even small, is observed between the first and the last stations of a cycle. This happens especially when cumulative inaccuracies are registered during the cycle survey without a final rectification by speleologists. This can be detected by a neighbourhood distance scanning around each point (Fig. 3 - case B).
- Missing connections are similar to cycle closure errors. This division of a junction survey station into two points separated by a small gap appears at junctions, notably when the explorations of the joining conduits were performed at different times and/or by different teams.

Moreover, survey stations are chosen by speleologists for their ease of access and clear sight along the cave passage. This choice has several consequences:

- Karstic networks are not regularly sampled. Additional samples can affect the apparent topology (Fig. 3 - case F) of the network. Some small meandering of the conduits could be ignored during the acquisition if a lower sampling resolution is chosen (Fig. 3 - case C).

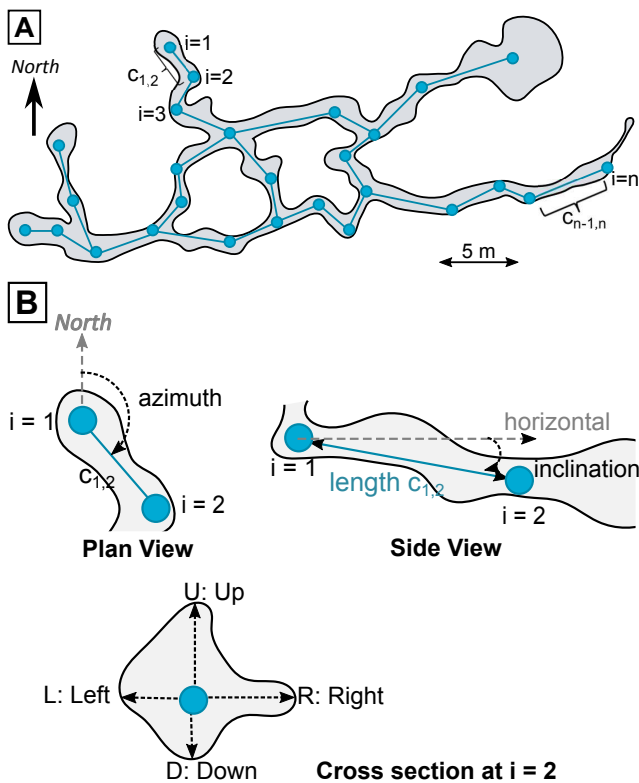


**Figure 1** The data set: 31 three-dimensional cave survey data have been gathered thanks to the help of various speleologists who agreed to share them. Three 2D networks were also used in the study and complete the database: Blue Spring, Crevice, and Crossroads. Four parts of the Sieben Hengste karst are also studied: SP1 and SP2 are subparts also included in the LargePart network; the UpPart is an independent one, not considered in the LargePart network. A relative altitude scale is indicated to ease the perception of the third dimension. The total cave survey length is also indicated. But for data protection reasons, we are not allowed to give more precise location, and scale of the networks.

**Table 1**

Karst dataset:  $L$  is the total mapped length (in kilometers);  $\Delta Z$  is the vertical extension of the system (in meters)

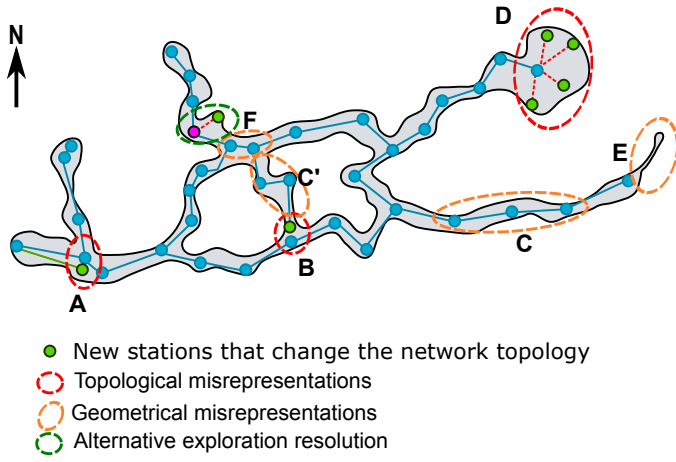
Name	Location	L (km)	$\Delta Z$ (m)	Name	Location	L (km)	$\Delta Z$ (m)
AgenAllwed	South Wales, UK	13.7	122	Kruber	Georgia	13.2	2191
ArphidiaRobinet	France	13.5	634	Lechuguilla	New Mexico, USA	329	1257
Arrestelia	France	60.9	827	Llangattwg	South Wales, UK	0.90	30
BlueSpring (2D)	Tennessee, USA	8.00	-	Mammuthöhle	Austria	63.9	1202
Ceberi	France	7.20	310	Monachou	France	0.70	33
CharentaisHeche	France	13.5	433	OjoDelAgua	Cuba	12.3	91
ClydachGorge Caf1	South Wales, UK	2.30	112	OxBelHa	Mexico	143	29
ClydachGorge OgofCapel	South Wales, UK	0.80	30	PicDuJer	France	0.60	67
Crevice (2D)	Missouri, USA	4.90	-	Ratasse	France	3.60	445
Crossroads (2D)	Virginia, USA	7.80	-	SaintMarcel	France	56.5	275
DarenCilau	South Wales, UK	20.2	186	Sakany	France	7.50	141
EglwysFaen	South Wales, UK	1.40	19	Shuanghe	Chine	130	593
FoussoubieEvent	France	2.60	130	SiebenHengsteLargePart	Switzerland	82.2	988
FoussoubieGoule	France	20.2	129	SiebenHengsteUpPart	Switzerland	3.30	131
GenieBraque	France	2.70	229	SiebenHengsteSP1	Switzerland	7.50	286
GrottesDuRoy	France	3.80	330	SiebenHengsteSP2	Switzerland	6.80	239
HanSurLesse	Belgium	9.80	30	Wakulla	Florida, USA	17.8	93



**Figure 2** Example of a cave map (plan view) and associated survey data. (A) 27 measurement stations  $i = 1, 2, \dots, n$  linked by 28 lines-of-sight  $c_{i,j}$  constitute the karst network. (B) Plan and side views: definition of length, azimuth, and inclination along a line-of-sight. Cross section: in the best case four measures (up, down, left, right) are made; more often, only width and height are recorded.

- Stations are not located on the *mathematical central line* of the conduit, i.e., they are not located at the middle of the conduit section. Placing the stations along the conduit walls can emphasize the conduit apparent sinuosity (Fig. 3 - case C’).
- Several stations can be set up in one single big cave in order to get a correct representation of its scale and to better map it. The result is either a cycle or some sun-ray shape of what should have been only registered as a point (Fig. 3 - case D).
- No genetic consideration is done for the definition of a line-of-sight, thus, owing to a change of the genetic phase, the form of the conduit can change radically between two survey stations without being exhaustively recorded and located.
- Finally, we cannot be sure of the completeness of the data: some conduits are not mapped because (i) they have not been explored yet, and/or (ii) they are not accessible, being too small or drowned on long distances (Fig. 3 - case E).

These approximations, of no consequence for speleological exploration, can affect the results of a systematic shape analysis. Two pre-processing steps have been performed to limit the survey errors. First, we have implemented a tool to correct the most common cases of missing connections. If an extremity vertex is closer than a user-defined tolerance distance to a neighbouring vertex or segment, it creates a new link between them. The case of two intermediate segments close to each other in 3D, and thus, possibly defining a junction, has been ignored: at this stage, we suppose that crossroads are important enough for speleologists to ensure that a station would have been defined on each real junction. This assumption has been verified on the 34 cave survey data we treated. Second, to deal with additional cycles that may have been generated by the previous treatment, we suppressed cycles smaller than a user-defined tolerance sphere. These corrections do not cover



**Figure 3** Errors linked to the data acquisition process. Cases A and B can be automatically detected and corrected when processing the data. (A) additionally point and line-of-sight result from a back-sight measurement; the green point and segment should not have been recorded. (B) a small gap appears between the first and last point of the cycle, resulting in a cycle closure error. C and C' - impact of the sampling location on the apparent sinuosity of the explored conduit: a regular sampling and systematically positioning of the station in the middle of the conduit could limit it. (D) sun-ray configuration resulting from the strategy chosen to sample one big cave. (E) some parts of the network remain unexplored owing to accessibility criteria. (F) the addition of a new station can change the local topology of the network.

all the cases mentioned above, but deals with the most automatically detectable errors.

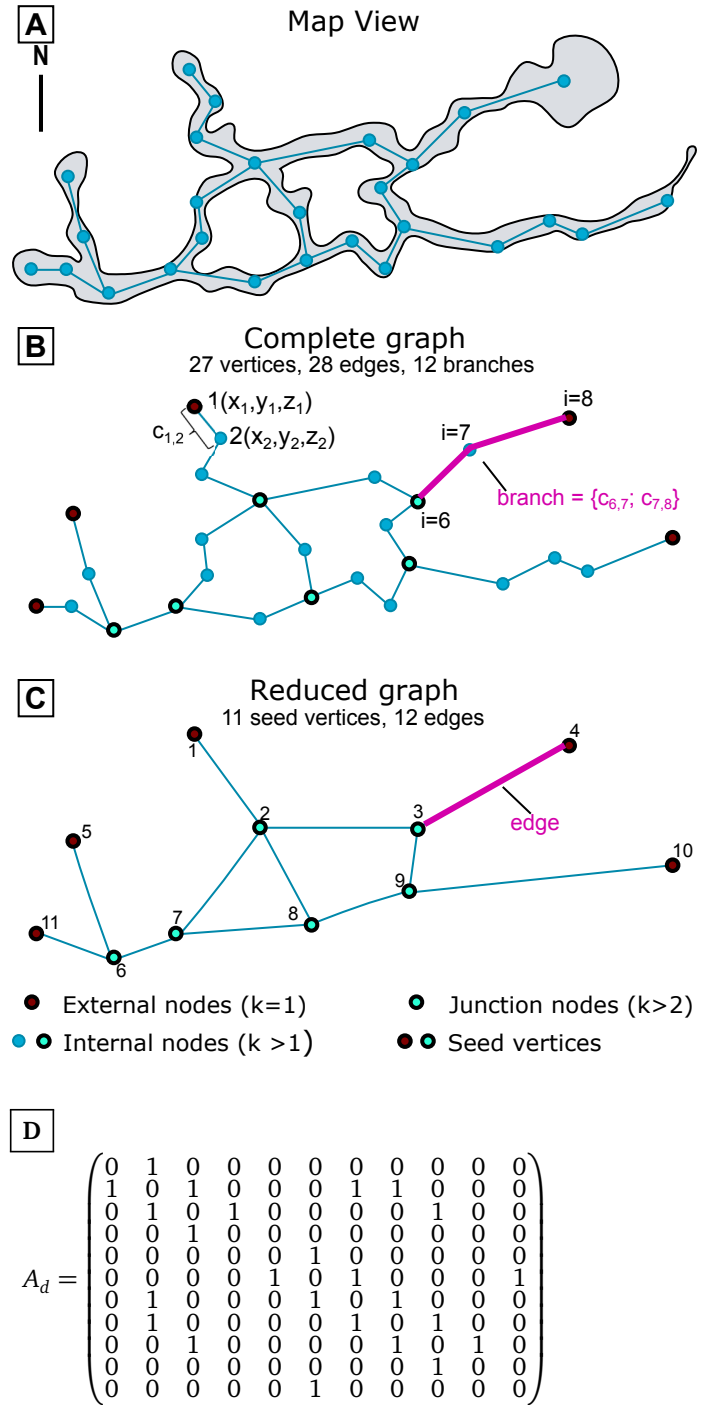
### Karst networks as graphs

Karst networks are considered in the analysis as a mathematical graph. This consideration is not new, as Howard [1971] already proposed such representation of natural systems to quantitatively analyse what was represented on planar maps, and several other authors have performed equally since [e.g., Glennon, 2001, Glennon and Groves, 2002, Pardo-Iguzquiza et al., 2011, Piccini, 2011].

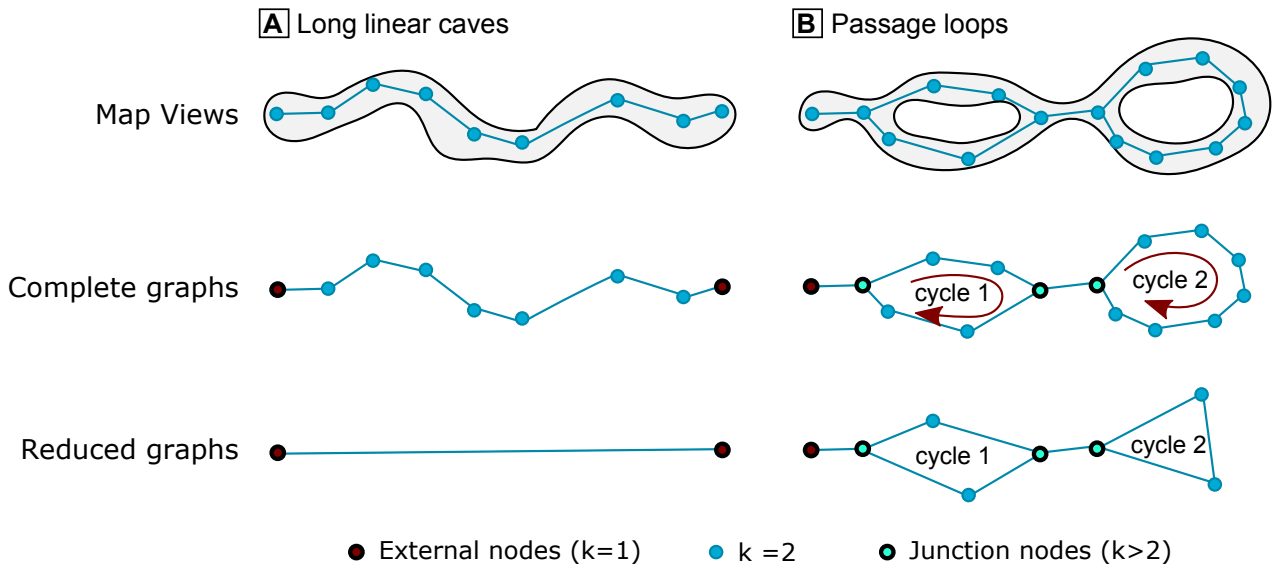
From the preprocessed networks, two different representations are used for the statistical analysis. For geometrical analysis, which requires real distance computations, the *complete networks* are directly used. But for topological parameters, a *reduced* representation has been defined that fastens the computations.

**Complete graph** In the complete graph of a karst, an edge  $c_{i,j}$ , or link, represents a line-of-sight between two stations that correspond to graph vertices  $i$  and  $j$  (Figs. 4A-B). Each vertex, or node, is characterised by its Cartesian coordinates  $i = \{x_i, y_i, z_i\}$ . The initial data in distance, direction, and inclination line-of-sight successions from an origin point have been converted to  $\{x, y, z\}$  coordinates to end with a network composed of a set of  $\mathcal{E}n$  vertices and  $s$  edges connecting them. In the karst perspective, the direction of a segment may intuitively be associated with the direction of the flow in the conduit. But this direction of flow is sometimes ambiguous, notably because of the presence of cycles in the karst network. Thus, the graph is undirected.

The *degree*  $k_i$  of a vertex is defined as the number of edges that are linked to this vertex. The *first neighbours* of a vertex



**Figure 4** From field data to graphs. (A) Map view showing the karst network and the survey stations. (B) Corresponding graph with 5 external nodes and 23 internal nodes, among which 6 are junction nodes; 28 edges are gathered into 12 branches. (C) Reduced graph: only the seed vertices are kept to obtain a topological simplified representation of the network. (D) Undirected adjacency matrix representation of the reduced graph.



**Figure 5** From field data to graphs: (A) Long linear caves; (B) Special cases where nodes of degree 2 are kept in the reduced graph to preserve the topology.

$i$  are the vertices that can be reached from  $i$  following one unique edge. Depending on the degree of a vertex we define (Fig. 4):

- the *extremity nodes* as the vertices with  $k_i = 1$ ;
- the *internal nodes* as the vertices with  $k_i > 1$ ; among them, the *junction nodes* are the vertices with  $k_i > 2$ .

A *branch* is defined as a set of adjacent edges connecting vertices of degree 2. They represent the portion of the curve connecting an extremity or a junction node with another junction or extremity node (Fig. 4B).

**Reduced graph** The number of internal nodes of degree  $k = 2$  in a topographical survey is largely a function of the field conditions and the speleologist's sampling preferences. These vertices do not give information about the topology of the karstic network. To simplify and speed up the topological analysis of the network, we defined a *reduced graph*. The reduction process consists in removing all vertices of degree 2 from the network (Figs. 4C and 5A). Thus, in the reduced graph, only junction and extremity nodes are kept, gathered in the term of *seed vertices*. Two seed vertices are then linked by an edge of unit length that replaces the branch of the corresponding explored path (Fig. 4C). The reduced graph is thus composed of  $N$  seed vertices and  $S$  edges linking them.

This process, while straightforward, takes also into account special cases. Indeed, when two different conduits link the same two seed vertices, forming a cycle, the cyclic structure has to be kept in the reduced network. To do so, one of the intermediate vertex is kept in each conduit (Fig. 5B: cycle 1). Another special situation comes from the cycles in the network. Cycles are defined as conduits starting and ending at the same seed vertex, without crossing any other seed vertex. In order to preserve the cyclic structure in the reduced network, two intermediate nodes are also kept in the looping conduit (Fig. 5B: cycle 2).

**Numerical representations** Two representations are used for the networks. The *position-links* or *node-links* representation corresponds to geometrical description of the network. It consists in two matrices: a matrix  $[X Y Z]$  storing the positions of the vertices and a matrix  $[i j]$  of edges between the vertices.

This representation can be visualized in 3D in the Gocad geo-modelling software or in Matlab, for whom specific functions were coded. For the real karstic networks, if the conduit diameter at each node is known, the position matrix can be extended with an additional column  $D$  leading to the new matrix  $[X Y Z D]$ .

To quickly and easily compute topological parameters, an adjacency matrix representation is also used (Fig. 4D). It is a square matrix  $A_d$ , where each element  $w_{ij}$  expresses the existence ( $w_{ij} = 1$ ) or absence ( $w_{ij} = 0$ ) of an edge from vertex  $i$  to vertex  $j$ . As we consider undirected networks, the adjacency matrix is symmetric as each edge from  $i$  to  $j$  is also an edge from  $j$  to  $i$ . In this work, we do not allow self-connecting vertices, i.e., vertices that have a link to themselves, and thus  $w(i, i) = 0 \forall i$ . Using the adjacency matrix, one can easily compute some statistics on the networks. The degree of each vertex is for instance the row-sum or the column-sum of the adjacency matrix.

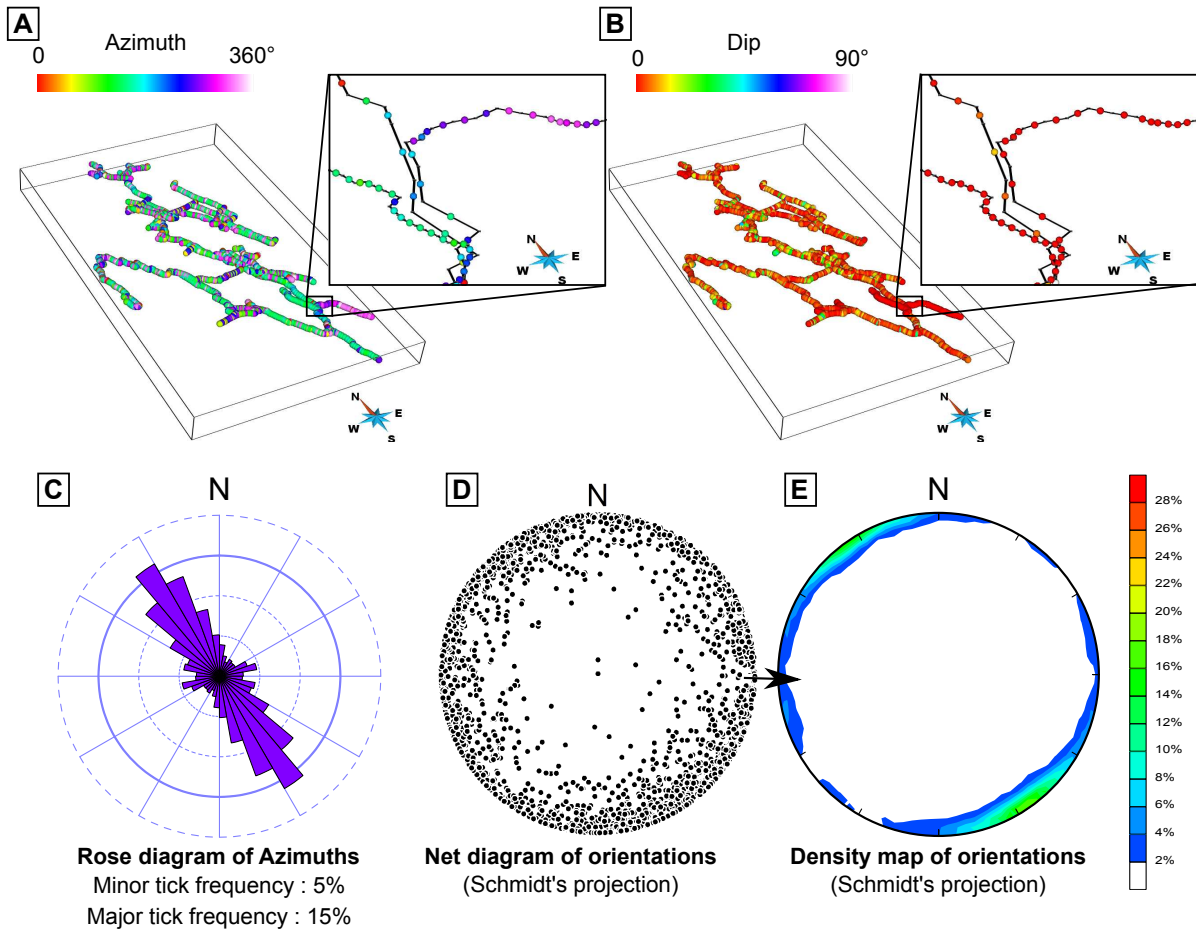
## METRICS TO CHARACTERIZE NETWORK GEOMETRIES

We propose here several metrics to characterize the karstic network geometries. They are all calculated on the complete graph of networks stored in a position-link representation. For each metric, we discuss the results obtained on the 31 three-dimensional networks of our database.

### Conduit orientation

Several geological features constitute natural drains for underground fluids and thus strongly influence the development of karstic conduits. Fractures count among those main features [Palmer, 1991]. Fractures are generally organized into families of particular orientation depending on the regional stress field [e.g., Billaux et al., 1989, Zoback, 1992, Beekman et al., 2000]. As a result, karstic networks that are mainly developed along the prominent fractures will show a *network pattern*, i.e., an angular grid of intersecting passages [Palmer, 1991].

To analyse the 3D network geometry, it is thus relevant and classical to compute the conduit orientations [e.g., Kiraly et al.,



**Figure 6** Orientation analysis of the Daren Cilau karst. (A) and (B) Azimuth (A) and dip (B) properties stored on the pointset corresponding to the midpoints of each edge ( $s = 2687$ ). (C) Rose diagram (symetric) showing a preferential orientation of the edges along the N160° (= N330°) direction. Frequency computation is weighted by the length of the edge projection on the horizontal plane to take into account the heterogeneous sampling step and the absence of direction associated to vertical conduits. The corresponding orientation entropy  $H_O = 0.921$  is among the lowest measured values, expressing the existence of a preferential direction. (D) Projection of the orientations on a Schmidt's stereonet (equal areas). (E) Density map computed on the Schmidt's stereonet (% of data in 1% of the entire stereonet area). The density map enhances the existence of a preferred direction family N160°-O° dip and of the close to horizontal development of this karstic network.

1971, Ford and Williams, 2007, Pardo-Iguzquiza et al., 2011]. In 3D, edges are assumed to point downward. Orientation of an edge consists in measuring its azimuth (angle between the north and the segment, ranging from 0 to 360°) and its dip (angle with the horizontal plane, ranging from 0 to 90°). Resulting values are stored on a 3D pointset corresponding to the midpoints of each line-of-sight (Figs. 6A and 6B).

Typically, these data are analysed with a Rose diagram to represent azimuth and dip distributions (Fig. 6C). No azimuth can be affected to vertical conduits, and edge lengths vary. Thus, to compute the Rose diagram, each orientation value has been weighted by the length of the edge projection on the horizontal plane.

To better assess the third dimension and combine azimuth-dip analysis, we also project the results on a Schmidt's stereonet (which preserves the areas) and draw a weighted density map (Figs. 6D and 6E). The density map counts the number of data points inside a cell counter that has by default a radius one tenth of the stereonet. For density map computation, to avoid bias owing to a heterogeneous sampling step, each point is weighted by the real edge length. The density map expresses so the percentage of orientation data that are contained in each 1% of the entire stereonet area. Data interpretation of the density map is done manually. We consider that preferential orientations exist when localised data clusters clearly

appear ( $density \geq 4\%$ ). Compared to the Rose diagram, the density map is essential to identify significant proportions of subvertical edges (Fig. 7). Families of poles can be defined for which circular statistics are provided. As the three 2D networks come from initial raster data, the orientation of their segments is aligned with the background grid making them irrelevant for such an analysis.

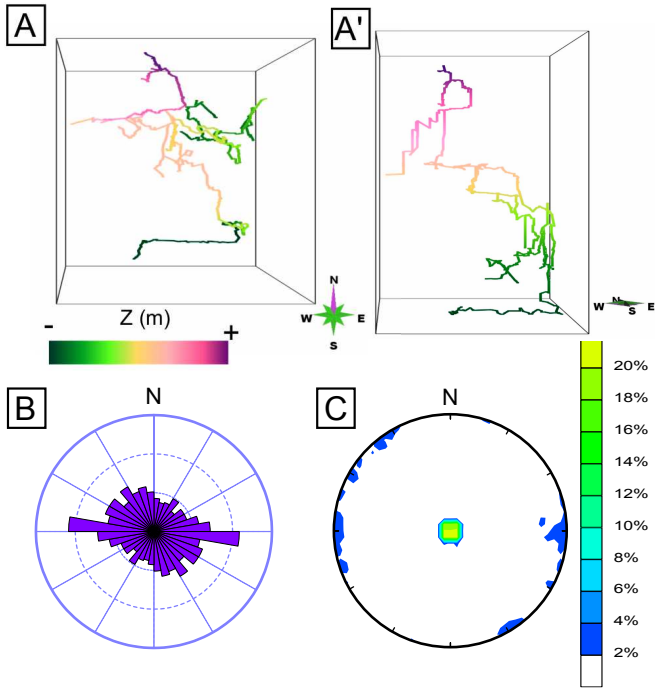
To better quantify the differences in orientations of conduits, we propose to compute the entropy of orientations,  $H_O$ , as recently done for urban street networks [Gudmundsson and Mohajeri, 2013]. Entropy is generally assimilated to a measure of disorder [e.g., Ziman, 1979, Journal and Deutsch, 1993]. We use the Shannon entropy formula:

$$H_O = - \sum_{i=1}^t p_i \ln_{nbins}(p_i) \quad (1)$$

with  $p_i$  the probability of an edge (weighted by its length) to fall in the  $i$ -th bin,  $nbins$  represents the number of bins when computing the histogram of the orientations and  $t$  the number of bins with nonzero probabilities. Note that the entropy calculation uses a base  $nbins$  logarithm.

As the orientation can only take values between 0° and 360°, and as we used undirected graphs, data of opposite directions are counted in the same bin and Rose diagrams are symmetric





**Figure 7** Orientation analysis of the Ratasse karst: an example of network with two preferential orientations: a subvertical one and a subhorizontal one. (A) and (A') Top and southern views of the Ratasse karst. (B) Rose diagram enhancing the moderate subhorizontal preferential orientation along N95-275° direction, which explains the orientation entropy  $H_O = 0.986$ . (C) Density map computed on the Schmidt's stereonet enhancing the clear preferential vertical orientation of some of the conduits (dip close to 90° gathers up to 20% of the edges) clearly visible on the southern view A'.

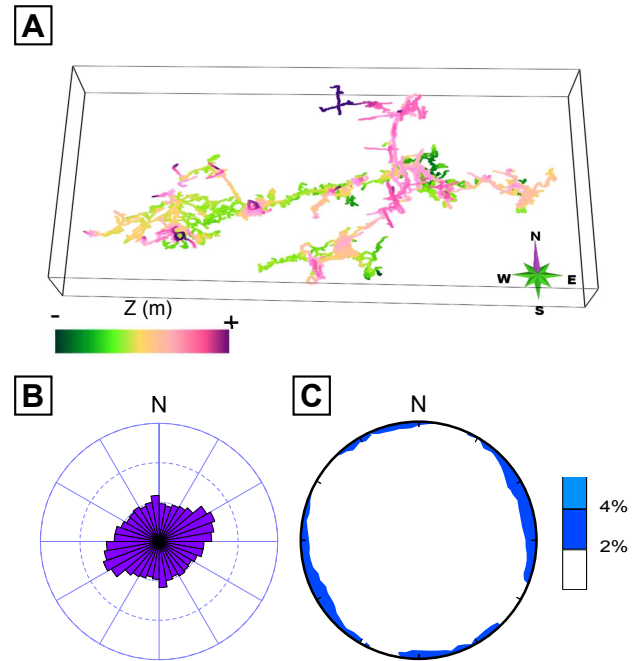
ones. This is also handled in the orientation entropy computation where we used bin widths of 10°, thus fixing the number of bins at 18. The shape of the probability distribution is thus quantitatively assessed by the entropy value that equals 0 if all edges occupy a single bin and that equals 1 if all bins have exactly the same probabilities (e.g., the distribution is uniform).

On the 31 three-dimensional karst networks that we analysed, the orientation entropy varies from 0.746 (for ClydachGorge Caf1) to 0.997 (OxBelHa). These globally high values express the fact that all directions are often observable. When preferred directions exist, their relative frequency is rarely superior to 20% of all data. In the data set indeed, the ClydachGorge Caf1 karst appears as an exception: it is a quite linear network, and the orientation analysis shows a maximal frequency of 32% along N150°, which explains the entropy value  $H_O = 0.794$ . Ranged by increasing entropy, the next network is thus Genie Braque with  $H_O = 0.884$  that corresponds to two main orientations, one clearly marked around the N50° direction. Then, the entropy of orientation increases regularly as preferred orientations are less and less distinguishable (A) until values are very close to 1. For example, the karst of Lechuguilla has an entropy of  $H_O = 0.996$ , which expresses the fact that all directions are almost equally observed in the network (Fig. 8). It corresponds to sinuous and curvilinear patterns developed when the passages are mostly influenced by almost horizontal bedding planes. Fixing a threshold value for which no preferential direction would be defined is quite difficult, but entropy of orientations appears as a good quantitative way to classify networks upon an orientation criteria. Nonetheless, caution has to be taken in the interpretation as

the entropy is only computed on the horizontal projection of the orientations. Thus, a preferential subvertical orientation does not appear on the Rose diagram and is ignored in the orientation entropy computation. This confirms the usefulness of the complementary density map analysis. In A, the karstic networks that show a preferred subvertical orientation are classified separately (Fig. 23): again from lowest orientation entropy (which is observed for networks with 2 preferred orientations, a subhorizontal and the vertical one: Arphidia Robinet,  $H_O = 0.918$ ) to the highest orientation entropy (observed for Krubera,  $H_O = 0.996$ , which is mainly vertically developed).

The variety of orientation analysis patterns observed on just 31 networks shows the variability that one can encounter when studying karstic systems. Most of the time, preferential orientations relate to particular inception features: tectonic (joints, fractures and faults) or stratigraphic (bedding planes: Filipponi et al. [2009]). For example, in the Sieben Hengste Large Part network, a main orientation of N100°/ 16° plunge is identified complementary to a vertical one. This is consistent with the dip and direction of the main karstifiable formation in which the system developed [Jeannin, 1996]. Also, the direction observed in the Han-sur-Lesse network (N95°/ 1° plunge) is linked to one preferential direction of fracturation observed in the field [Bonniver, 2011].

Conduit orientation is thus an interesting parameter for detecting the geological features of influence and for better understanding the speleogenetic processes that have locally dominated. In this way, entropy of orientations constitutes a useful metric to quantitatively assess the existence and relative importance of preferential karstic developments.



**Figure 8** Orientation analysis of the Lechuguilla karst: an example of a network with no preferential orientation. (A) Top view of the Lechuguilla karst. (B) and (C) Rose diagram and density map showing the absence of preferential orientation conduit, confirmed by the very high value of orientation entropy  $H_O = 0.996$ .

**Table 2**

Geometry measures on the 31 three-dimensional karst networks ( $L$  is the total explored length (in kilometers);  $H_O$  is the entropy of orientations;  $n$  is the number of survey stations, i.e., the number of nodes of the complete graph;  $s$  is the number of line-of-sights, i.e., edges;  $\overline{len}$  is the average length of karst branches (in meters);  $H_{len}$  is the entropy of karst branch lengths;  $CV_{len}$  is the coefficient of variation of the branch lengths, expressed in %;  $\bar{t}$  is the average tortuosity of karst branches); for each parameter, minimal and maximal values are indicated in **bold**

	$L$ (km)	$H_O$	$n$	$s$	$\overline{len}$ (m)	$H_{len}$	$CV_{len}$ (%)	$\bar{t}$
AgenAllwed	13.7	0.964	1874	1893	119.99	0.46	1.84	<b>1.54</b>
ArphidiaRobinet	13.5	0.918	1461	1477	70.73	0.63	1.09	1.42
Arrestelia	60.9	0.907	6257	6268	69.74	0.34	1.49	1.35
Ceberi	7.2	0.972	881	886	55.61	0.50	1.35	1.27
CharentaisHeche	13.5	0.986	1985	2013	50.84	0.22	<b>2.56</b>	1.25
ClydachGorge Caf1	2.3	<b>0.746</b>	230	229	<b>331.74</b>	0.55	1.51	1.18
ClydachGorge OgofCapel	<b>0.8</b>	0.892	145	146	50.80	0.58	1.55	1.35
DarenCilau	20.2	0.921	2665	2687	55.85	0.32	1.87	1.21
EglwysFaen	1.4	0.948	307	308	15.81	0.68	1.29	1.14
FoussoubieEvent	2.6	0.928	333	339	33.20	0.50	1.61	1.27
FoussoubieGoule	20.2	0.987	2326	2354	62.30	0.66	1.41	1.21
GenieBraque	2.7	0.884	164	163	194.03	<b>0.74</b>	0.95	1.20
GrottesDuRoy	3.8	0.981	724	733	29.76	0.65	1.10	1.24
HanSurLesse	9.8	0.976	1668	1705	62.06	0.63	1.26	<b>1.12</b>
Kruber	13.2	0.996	2150	2157	55.31	0.53	1.87	1.30
Lechuguilla	<b>329</b>	0.996	12725	13503	68.96	0.57	0.91	1.26
Llangattwg	0.9	0.954	232	228	15.32	0.66	1.14	1.15
Mammuthöhle	63.9	0.996	9348	9712	30.99	0.25	1.29	1.42
Monachou	0.7	0.984	251	258	<b>8.46</b>	0.59	<b>0.90</b>	1.24
OjoDelAgua	12.3	0.964	1292	1326	43.81	0.53	1.43	1.21
OxBelHa	143	<b>0.997</b>	10098	10098	170.79	0.65	0.94	1.37
PicDuJer	0.6	0.933	102	102	25.51	0.64	1.12	1.21
Ratasse	3.6	0.986	692	693	64.86	0.44	1.59	1.41
SaintMarcel	56.5	0.993	5506	5556	75.96	0.40	1.78	1.23
Sakany	7.5	0.992	1716	1784	20.81	0.49	0.92	1.40
Shuanghe	130	0.988	7581	7634	118.76	0.46	1.28	1.22
SiebenHengsteLargePart	82.2	0.988	15340	15570	36.84	<b>0.18</b>	1.73	1.36
SiebenHengsteUpPart	3.3	0.978	753	753	52.46	0.72	1.04	1.36
SiebenHengsteSP1	7.5	0.991	1881	1915	26.94	0.58	1.32	1.40
SiebenHengsteSP2	6.8	0.991	1396	1418	39.52	0.64	1.21	1.43
Wakulla	17.8	0.982	474	477	330.51	0.56	1.53	1.28

## Length, length entropy, and coefficient of variation of the lengths

The curvilinear length of a branch  $l_{i,j}$  is measured from one seed vertex  $i$  to the following one  $j$  by adding the length of each edge it contains. The average branch length varies from 8.46 m for the Monachou cave to 331.74 m for the Clydach Gorge Caf 1 karstic system (Table 2). This said, most of the values range from 20 to 70 m with a set of only six networks showing higher values: Agen Allwed, Clydach Gorge Caf1, Genie Braque, Ox Bel Ha, Shuanghe, and Wakulla. Those networks do not correspond to a specific pattern: the first three are elongated with one main long conduit, but the three others are more branchwork ones. One could have supposed that exploration in the largest caves could have been done with a looser sampling, inducing a smaller average branch length. This is not at all the case, as no linear relationship appears with the total survey length (linear correlation coefficient  $r = 0.07$ ).

As for orientation, we also propose to compute the Shannon entropy of the lengths  $H_{len}$  to measure the variability of conduit lengths in the network: length entropy will be maximal for a uniform distribution and equal to zero if all conduits have

the exact same length. In practice, contrary to orientations, lengths are not limited values. To have comparable entropy values between all networks, branch lengths are normalized by the maximum length of the network. Then, results are binned in 10 bins. In this study, the entropy of the lengths is therefore defined as

$$H_{len} = - \sum_{i=1}^t p_i \log_{10}(p_i) \quad (2)$$

with  $p_i$  the weighted probability of branches falling in the  $i$ -th bin, and  $t$  is the number of bins with nonzero probabilities of edges.

We also computed the coefficient of variation of the lengths  $CV_{len}$  to characterize the dispersion of the measures. It is expressed in percentage and corresponds to the ratio of the standard deviation  $\sigma_{len}$  to the mean  $\overline{len}$ :

$$CV_{len} = \frac{\sigma_{len}}{\overline{len}} * 100 \quad (3)$$

The length entropy  $H_{len}$  varies from 0.18 to 0.74. The coefficient of variation of the lengths  $CV_{len}$  varies from 0.90% to 2.56% (Table 2). No clear linear relationship is observable between both parameters, but a slight negative correlation is

observed ( $r^2 = -0.69$ , see also section 5). The Charentais Heche network is a good example to better illustrate the respective meaning of both these parameters (Fig. 9). This network has the highest coefficient of variation of our data set,  $CV_{len} = 2.56\%$ : branch lengths vary indeed from low (around 2 m) to high values (around 1550 m) for an average length  $\overline{len} = 50$  m. But in the case of Charentais Heche, these extrema are *outliers* that squeeze the data into the first three bins (Fig. 9), inducing a low entropy of lengths,  $H_{len} = 0.18$ . On the opposite, the Genie Braque karst is characterized by the highest length entropy of our data set ( $H_{len} = 0.74$ ), which expresses the more uniform distribution of the lengths (Fig. 9). In this case, the branch lengths do not vary very much around the mean, which is expressed by a very low value of  $CV_{len} = 0.95\%$ .

## Conduit tortuosity (or sinuosity)

Tortuosity, also called sinuosity, is a classical metric used to characterize karstic morphologies [Jeannin et al., 2007, Pardo-Iguzquiza et al., 2011]. It is used as well to parameterize karst network simulations [Pardo-Iguzquiza et al., 2012]. It is addressed at the scale of a branch. The tortuosity  $t_{i,j}$  from one seed vertex  $i$  to the following one  $j$  is the ratio between the curvilinear length  $l_{i,j}$  along the branch and the euclidean length  $d_{i,j}$  between the two seed vertices:

$$t_{i,j} = \frac{l_{i,j}}{d_{i,j}} \quad (4)$$

In order to avoid divergence for the tortuosity coefficient, we discard all looping path from the computation. The tortuosity coefficient of a network is defined as the mean tortuosity  $\bar{t}$  of all the branches in the network.

Tortuosity coefficients range from 1.12 (Han-sur-Lesse) to 1.54 (Agen Allwed: Table 2). Tortuosity values are quite difficult to interpret. Indeed, one could have assumed that high tortuosity coefficients would reflect sinuous, curvilinear patterns of conduits like branchwork or anastomotic caves as defined by Palmer [1991]. But the results seem more linked to the sampling strategy than to a clearly cave pattern (Fig. 10). This refers to the limitations of the data acquisition process, which is not guided by statistical considerations but by surveying constraints (section 2.1), which also vary depending on the speleologist team. Nevertheless, this parameter is sometimes proposed as an input parameter for stochastic simulations [Pardo-Iguzquiza et al., 2012].

## STATISTICAL MEASURES OF TOPOLOGY

To complement the geometrical analysis of karstic networks, we propose to use several parameters from graph theory to characterize network topology. Table 3 summarizes these metrics, which are computed on the reduced graph representations. As only the topology matters here, the three 2D networks are also included in the database, allowing us to have 34 networks to analyse.

### Considering cycles

In the following, *tree graphs*, or also called *acyclic networks*, refer to networks that do not have any cycles or passage loops. They correspond to branchwork patterns [Palmer, 1991]. The

**Table 3**

List of metrics used to characterize the karst topology; all are computed on the reduced graphs

$N_{cycl}$	Cyclomatic number (number of passage loops)
$D_C$	Connectivity degree
$\Theta$	Global cyclic coefficient [Kim and Kim, 2005]
$\bar{k}$	Average vertex degree
$CV_k$	Coefficient of variation of degrees
$r_k$	Correlation of vertex degrees [Newman, 2002]
$\overline{SPL}$	Average shortest path length
$CPD$	Central point dominance [Freeman, 1977]

terms *interconnected networks* or *maze patterns* are used to name *nonacyclic* networks, i.e. networks with several passage loops. These networks regroup the anastomotic (dominated by curvilinear conduits) and network patterns (dominated by straight conduits linked to the enlargement of fractures) of Palmer’s classification.

### Connected components and cyclomatic number

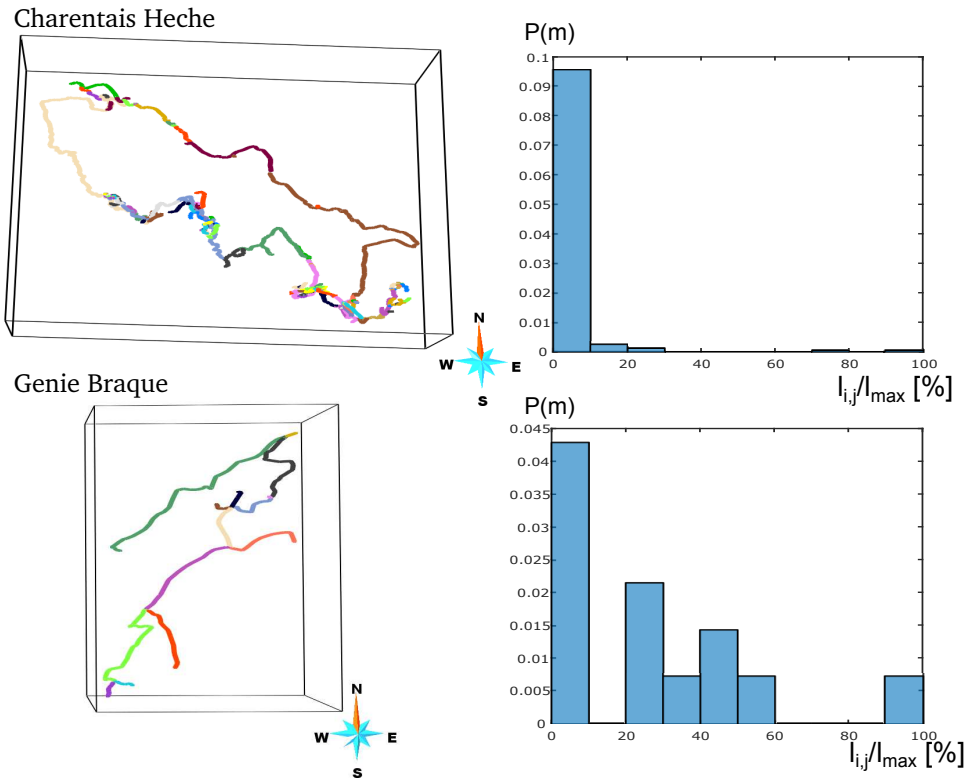
A connected component of a graph is a subgraph such that all its nodes are reachable by all other nodes in the subgraph. Reachability corresponds to the existence of a path between the nodes. To compute the connected components of a cave system we use a depth-first-search algorithm tagging each vertex with the index of the connected component to which it belongs (Fig. 11). The number of connected components corresponds to the number of subgraphs  $p$  as defined in graph theory. Our data set gathers karstic systems mostly composed by 1 to 5 connected components (Table 4). The Mammuthöhle karst has, however, the particularity of being composed by 25 connected components. Showing multiple connected components means that the network is divided into several parts that are not linked one to another by a mapped karstic conduit. Two possible interpretations can explain this absence of link (i) a hydrological connection does exist and has been proved by tracer tests, but the corresponding conduit(s) has(ve) not been yet, or cannot be, explored or (ii) the connection does not exist.

The cyclomatic number  $N_{cycl}$  is a classical parameter used in graph theory. It reflects the number of cycles, or passage loops, inside a network, which is high for maze patterns and equals 0 for acyclic networks. It can be automatically calculated from the total number of nodes  $N$ , the number of edges  $S$ , and the number of subgraphs  $p$ :  $N_{cycl} = S - N + p$ . A large variety of cases are encountered in our database, with cyclomatic numbers ranging from 0 to 779 for the highly anastomotic Lechuguilla karst.

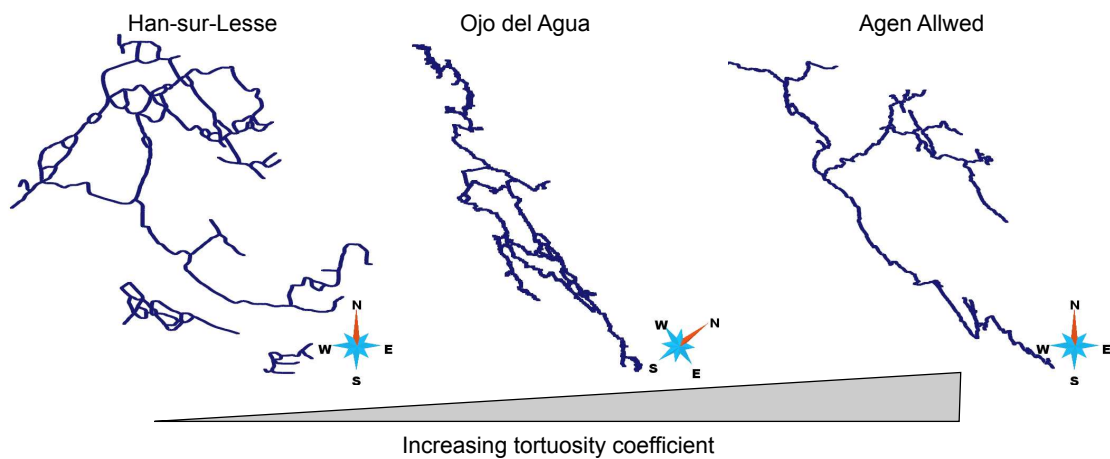
### Howard parameters

Derived from graph theory [Garrison and Marble, 1962, Kinsky, 1963], Howard’s parameters were developed to characterize braided patterns of streams [Howard et al., 1970]. They were rapidly applied to quantify karst network connectivity [Howard, 1971]. Howard’s parameters are defined for planar graphs and depend on three graph attributes: the number of external nodes  $N_{ext}$ , the number of junction nodes  $N_{junc}$ , and the cyclomatic number  $N_{cycl}$  (Fig. 4).

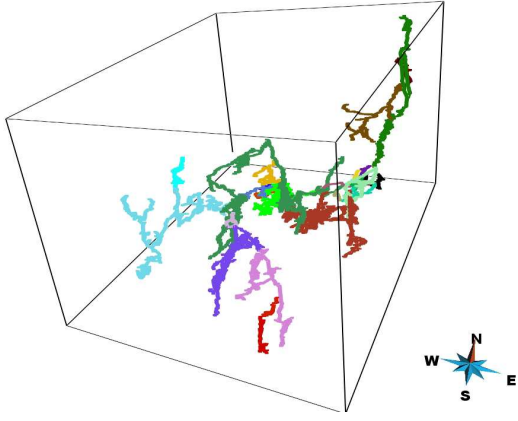
The parameter  $\alpha$  designates the ‘ratio of the observed number of cycles to the greatest possible number of [cycles] for a



**Figure 9** Charentais Heche and Genie Braque karsts: colors vary to enhance the different karst branches. On the right the corresponding length distributions are represented. The rather wide distribution of Genie Braque branch lengths is expressed through a high  $H_{len}$  value ( $H_{len} = 0.74$ ), while the more concentrated distribution of Charentais Heche corresponds to a low value of  $H_{len} = 0.18$ .



**Figure 10** Three top views of karst networks of increasing tortuosity coefficient: Han-sur-Lesse  $\bar{t} = 1.12$ , Ojo del Agua  $\bar{t} = 1.21$ , and Agen Allwed  $\bar{t} = 1.54$ . No evidence of a link between the cave pattern and the tortuosity coefficient arises.



**Figure 11** The 25 connected components of the Mammuthöhle network: each color represents one connected component.

given number of nodes' [Howard et al., 1970]:

$$\alpha = \frac{N_{cycl}}{2(N_{junc} + N_{ext}) - 5} \quad (5)$$

The parameter  $\beta$  is the ratio of the number of edges to the number of seed vertices:

$$\beta = \frac{S}{N_{junc} + N_{ext}} \quad (6)$$

The parameter  $\gamma$  expresses 'the ratio of the observed number of edges to the greatest possible number of edges for a given number of nodes' [Howard et al., 1970], i.e., the ratio between the number of karst branches toward the maximal possible connections:

$$\gamma = \frac{S}{3(N_{junc} + N_{ext} - 2)} \quad (7)$$

It is important to emphasize again that these parameters were developed for 2D planar graphs. In our data set, most networks are 3D, and their topology cannot always be reduced to those of their 2D horizontal projection (such as Arphidia Robinet, Mammuthöhle, or Krubera).

However, even if the assumptions underlying Howard's calculations are not met, it is interesting to compute those parameters in order to compare our results with the 25 networks that Howard studied. Howard [1971] indicates that  $\alpha$ ,  $\beta$ , and  $\gamma$  respective values are close to 0, 1, and 0.33 for a branchwork karst; and respectively close to 0.25, 1.5, and 0.5 for a reticular one. They can be thus synthesized into the connectivity degree  $D_C$ , expressed in percentage, which is expected to be close to 0 for a branchwork karst and close to 1 (or 100 if expressed in %) for a reticular karst:

$$D_C = \frac{\frac{\alpha}{0.25} + \frac{\beta-1}{0.5} + \frac{\gamma-0.33}{0.17}}{3} \quad (8)$$

The results are presented in Table 4. The  $\alpha$  values range from 0 to 0.16,  $\beta$  values from 0.88 to 1.28, and  $\gamma$  values from 0.32 to 0.44, which is quite consistent with the reference values for karst. High values of the three parameters systematically concern the same networks: Han-sur-Lesse, Sakany, Mammuthöhle, Sieben Hengste SP2. But for the low values the classification slightly varies depending on the considered parameters. In every case, low values correspond to obviously branchwork karsts, while the high values are associated to

more reticular patterns (Fig. 12). Between these extrema, all intermediate configurations (and values) are observed, and defining thresholds that would permit us to differentiate karst categories is quite difficult.

### Cyclic coefficient

The cyclic coefficient is a topological coefficient defined by Kim and Kim [2005] in order to measure how cyclic a network is. The local cyclic coefficient  $\Theta_i$  of a seed vertex  $i$  is defined as the mean of the inverse of the sizes of the smallest cycles formed by vertex  $i$  and its neighbours (Fig. 13A). The mean is taken on all possible pairs of edges connected to the vertex  $i$ :

$$\Theta_i = \frac{2}{k_i(k_i - 1)} \sum_{(j,h)} \frac{1}{L_{jh}^i} \quad (9)$$

where  $L_{jh}^i$  is the size of the smallest cycle that passes through vertices  $i$ ,  $j$ , and  $h$ . If no cycle passes through  $i$ ,  $j$ , and  $h$  then  $L_{jh}^i = \infty$  (Fig. 13A).

The algorithm for the computation of the cyclic coefficient for vertex  $i$  consists in taking its two first neighbours ( $j$  and  $h$ ) and removing the edges  $(i, j)$ ,  $(i, h)$  and their opposite counterparts  $(j, i)$  and  $(h, i)$  from the network. Then, the length  $L_{jh}$  of the shortest path between  $j$  and  $h$  is computed. The length of the cycle is then obtained by adding the removed edges  $L_{ijh} = L_{jh} + 2$ . These edges are then reintroduced in the network and the computation is redone on a second pair of neighbours of vertex  $i$  until each possible pairs are visited.

The global cyclic coefficient for a network is defined as the mean cyclic coefficient of all the seed vertices of the network:

$$\Theta = \frac{1}{N} \sum_i \Theta_i \quad (10)$$

If a network has a pure branchwork pattern, no cycle is present in the network,  $\theta = 0$  (Fig. 13B). On the opposite, if all vertices of a network are connected to the other ones, the cyclic coefficient will be equal to one-third (Fig. 13C). Note that the cyclic coefficient is slightly different from the global clustering coefficient,  $C$ , defined by Watts and Strogatz [1998], that is also used to measure how well a network is connected on a local neighbour-to-neighbour scale [e.g., Newman, 2003, Andresen et al., 2013].

The results obtained on the 34 karstic systems range from 0 (Clydach Gorge Caf 1 and Genie Braque) to 0.12 (Han-sur-Lesse) with a mean value of 0.04 (Table 4). An evident correlation appears with the connectivity degree and is further discussed in section 5.

### Measures on vertex degrees

#### Distribution and coefficient of variation of vertex degrees

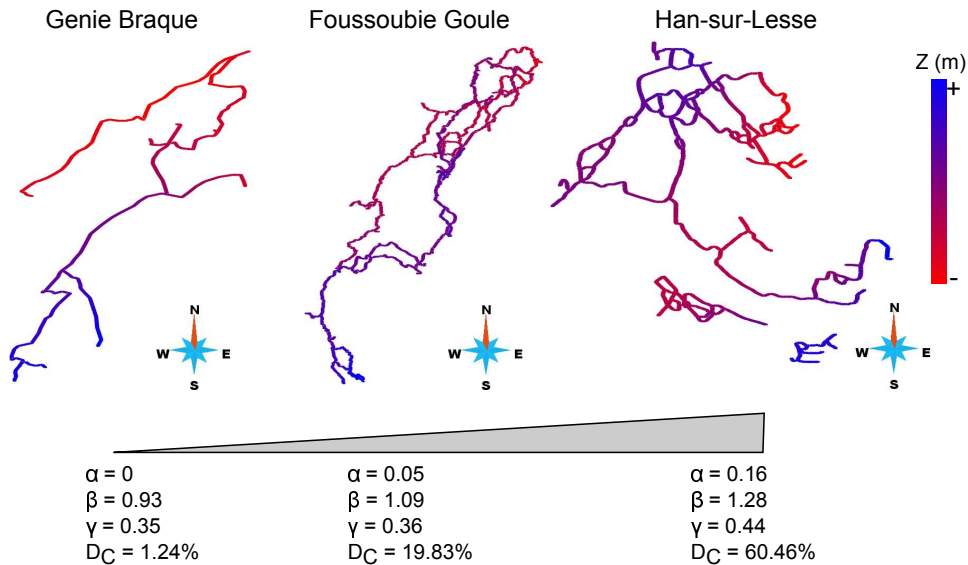
Simple statistics related with the vertex degrees can also be computed. These include the average vertex degree  $\bar{k}$  and the standard deviation of vertex degrees  $\sigma_k$ .

With a mean value of 2.14, the average vertex degree ranges from 1.75 to 2.57 on the 34 studied karsts (Table 5 and Fig. 15). To interpret these values, it is important to remember that this metric is calculated on the reduced graphs of the networks so that all nodes of degree 2 have been removed. Starting from this observation, if the studied graph is acyclic, i.e., no cycle is observable, and we consider a network with  $k_{max} = 3$ , a

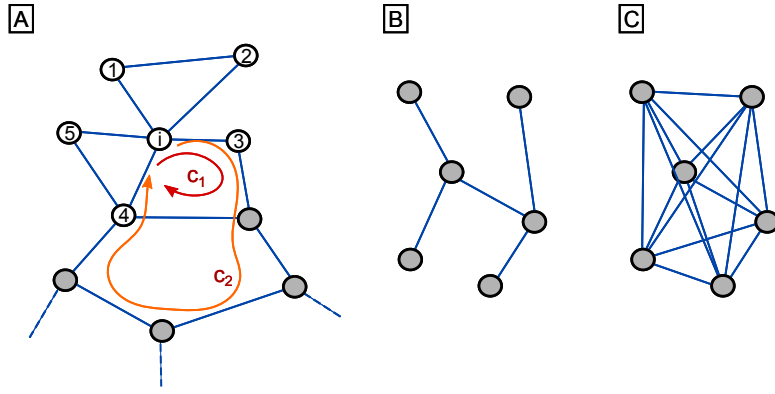
**Table 4**

Topology measures on the 34 reduced graphs of the karst networks: considering cycles ( $N$  is the number of seed vertices;  $S$  is the number of edges;  $p$  is the number of connected components;  $N_{cycl}$  is the cyclomatic number (number of passage loops);  $\alpha$ ,  $\beta$ , and  $\gamma$  are the Howard's parameters;  $D_C$  is the connectivity degree;  $\Theta$  is the global cyclic coefficient); maximal and minimal values are enhanced in **bold**

	$N$	$S$	$p$	$N_{cycl}$	$\alpha$ ( $10^{-2}$ )	$\beta$	$\gamma$ ( $10^{-2}$ )	$D_C$ (%)	$\Theta$ ( $10^{-2}$ )
AgenAllwed	129	148	1	20	7.91	1.15	38.85	31.82	5.70
ArphidiaRobinet	180	196	2	18	5.07	1.09	36.70	19.95	4.02
Arrestelia	879	890	3	14	0.80	1.01	33.83	3.52	0.83
BlueSpring (2D)	477	503	1	27	2.85	1.05	35.30	11.93	2.00
Ceberi	132	137	4	9	3.47	1.04	35.13	11.33	3.28
CharentaisHeche	260	288	1	29	5.63	1.11	37.21	22.94	5.17
ClydachGorge Caf1	8	7	1	<b>0</b>	<b>0.00</b>	<b>0.88</b>	38.89	3.21	<b>0.00</b>
ClydachGorge Ogof Capel	19	20	1	2	6.06	1.05	39.22	23.78	5.26
Crevice (2D)	253	262	1	10	2.00	1.04	34.79	8.55	1.55
Crossroads (2D)	371	430	1	60	8.14	1.16	38.84	32.92	5.91
DarenCilau	351	373	1	23	3.30	1.06	35.63	13.73	2.19
EglwysFaen	88	89	1	2	1.17	1.01	34.50	5.25	0.61
FoussoubieEvent	75	81	1	7	4.83	1.08	36.99	19.59	3.50
FoussoubieGoule	302	330	1	29	4.84	1.09	36.67	19.83	3.79
GenieBraque	15	14	1	<b>0</b>	<b>0.00</b>	0.93	35.90	1.24	<b>0.00</b>
GrottesDuRoy	124	133	2	11	4.53	1.07	36.34	17.42	3.58
HanSurLesse	130	167	3	40	<b>15.6</b>	<b>1.28</b>	<b>43.49</b>	<b>60.46</b>	<b>12.4</b>
Krubera	235	242	1	8	1.72	1.03	34.62	7.46	1.12
Lechuguilla	4187	4965	1	<b>779</b>	9.31	1.19	39.55	37.63	6.26
Llangattwg	67	63	5	1	0.78	0.94	<b>32.31</b>	<b>-4.30</b>	0.50
Mammuthöhle	2168	2532	25	389	8.98	1.17	38.97	34.87	6.44
Monachou	80	87	1	8	5.16	1.09	37.18	20.91	3.63
OjoDelAgua	258	292	1	35	6.85	1.13	38.02	27.76	5.10
OxBelHa	838	838	1	1	0.06	1.00	33.41	0.89	0.04
PicDuJer	26	26	1	18	2.13	1.00	36.11	7.58	2.56
Ratasse	59	60	1	2	1.77	1.02	35.09	7.58	1.13
SaintMarcel	713	763	3	53	3.73	1.07	35.77	15.08	2.83
Sakany	344	412	1	69	10.1	1.20	40.16	40.68	6.46
Shuanghe	1058	1111	3	56	2.65	1.05	35.07	10.93	1.91
SiebenHengsteLargePart	2063	2293	1	231	5.61	1.11	37.09	22.92	4.34
SiebenHengsteUpPart	65	65	5	5	4.00	1.00	34.39	8.06	3.24
SiebenHengsteSP1	250	284	1	35	7.07	1.14	38.17	28.64	5.24
SiebenHengsteSP2	156	178	1	23	7.49	1.14	38.53	30.23	6.03
Wakulla	55	58	1	4	3.81	1.05	36.48	15.54	3.56



**Figure 12** Three top views of karstic networks that have increasing values of Howard's parameters and connectivity degree: low values characterize branchwork patterns, while high values are observed for reticular systems.

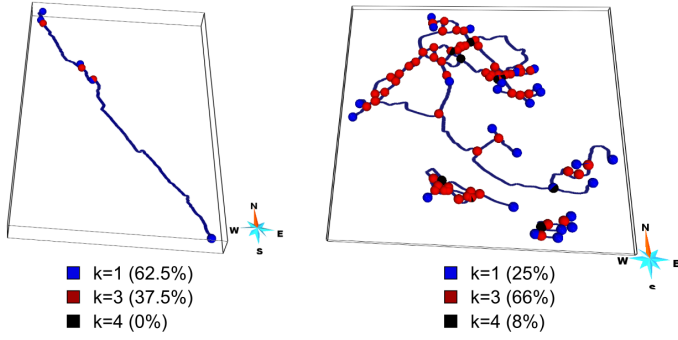


**Figure 13** Cyclic coefficient. (A) An example for the local cyclic coefficient computation: the vertex  $i$  has five neighbours,  $k_i = 5$ . Only one cycle passes through 1 and 2 or 4 and 5,  $L_{12}^i = L_{45}^i = 3$ . Two cycles pass through 3 and 5: by  $c_1$  and by  $c_2$  (four through 3 and 4). The shortest one is  $c_1$ , so  $L_{34}^i = 4$  and  $L_{35}^i = 5$ . Vertices 1 and 2 are not directly linked to vertices 3, 4 and 5,  $L_{13}^i = L_{14}^i = L_{15}^i = L_{23}^i = L_{24}^i = L_{25}^i = \infty$ . As a result,  $\Theta_i = 0.11$ . (B) and (C) Two examples of global cyclic coefficient values on a small network with  $N = 6$  vertices showing extrema cases: in case (B), no cycle is present in the network,  $\theta = 0$ ; in case (C), each vertex is always directly linked to the five other vertices, and all cycles have the smallest possible dimension of 3, so  $\theta = 0.33$

**Table 5**

Topology measures on the 34 reduced graphs of the karst networks: other parameters ( $\bar{k}$  is the average vertex degree;  $CV_k$  is the coefficient of variation of degrees;  $r_k$  is the correlation of vertex degrees;  $\overline{SPL}$  is the average shortest path length;  $CPD$  is the central point dominance); maximal and minimal values are enhanced in **bold**

	$\bar{k}$	$CV_k$ (%)	$r_k$	$\overline{SPL}$	$CPD$
AgenAllwed	2.29	50.05	0.63	11.72	0.54
ArphidiaRobinet	2.18	49.06	0.74	13.28	0.45
Arrestelia	2.03	57.14	0.77	29.16	0.50
BlueSpring (2D)	2.11	47.13	0.73	26.90	0.46
Ceberi	2.08	51.42	0.63	7.20	0.30
CharentaisHeche	2.22	45.91	0.69	25.40	0.43
ClydachGorge Caf1	<b>1.75</b>	59.15	<b>0.44</b>	<b>2.32</b>	0.56
ClydachGorge Ogof Capel	2.11	52.26	0.49	3.95	0.47
Crevice (2D)	2.07	48.35	0.70	19.66	0.55
Crossroads (2D)	2.32	41.75	0.83	17.33	0.34
DarenCilau	2.13	55.47	0.74	18.44	0.51
EglwysFaen	2.02	59.49	0.59	8.69	<b>0.57</b>
FoussoubieEvent	2.16	48.16	0.72	7.87	0.47
FoussoubieGoule	2.19	45.96	0.76	25.07	0.44
GenieBraque	1.87	<b>60.29</b>	0.55	3.37	0.51
GrottesDuRoy	2.15	48.91	0.75	11.36	0.31
HanSurLesse	<b>2.57</b>	<b>35.26</b>	0.85	7.14	0.20
Kruber	2.06	58.80	0.58	21.48	0.45
Lechuguilla	2.37	52.01	0.84	<b>55.76</b>	0.49
Llangattwg	1.88	59.65	0.52	4.05	0.16
Mammuthöhle	2.34	53.59	0.72	14.89	<b>0.02</b>
Monachou	2.18	50.03	0.75	7.58	0.40
OjoDelAgua	2.26	47.42	0.78	19.58	0.38
OxBelHa	2.00	51.79	0.65	49.42	0.54
PicDuJer	2.00	49.00	0.59	5.80	0.37
Ratasse	2.03	55.51	0.70	8.40	0.44
SaintMarcel	2.14	51.53	0.72	21.69	0.23
Sakany	2.40	47.08	<b>0.88</b>	12.99	0.27
Shuanghe	2.10	51.02	0.73	25.98	0.27
SiebenHengsteLargePart	2.22	48.12	0.77	47.37	0.46
SiebenHengsteUpPart	2.00	53.03	0.66	4.28	0.14
SiebenHengsteSP1	2.27	45.96	0.75	15.53	0.49
SiebenHengsteSP2	2.28	46.91	0.79	14.04	0.39
Wakulla	2.11	47.13	0.64	8.11	0.43



**Figure 15** Examples of karstic networks and corresponding average vertex degree: on the left, the Clydach Gorge Caf1, which is a tree network:  $\bar{k} = 1.75$ ; on the right Han-sur-Lesse, which has a large number of cycles and nodes of degree 4:  $\bar{k} = 2.57$ .

relation exists between the total number of nodes  $N$  and  $n_3$  the number of nodes of degree 3:

$$N = 2n_3 + 2 \quad (11)$$

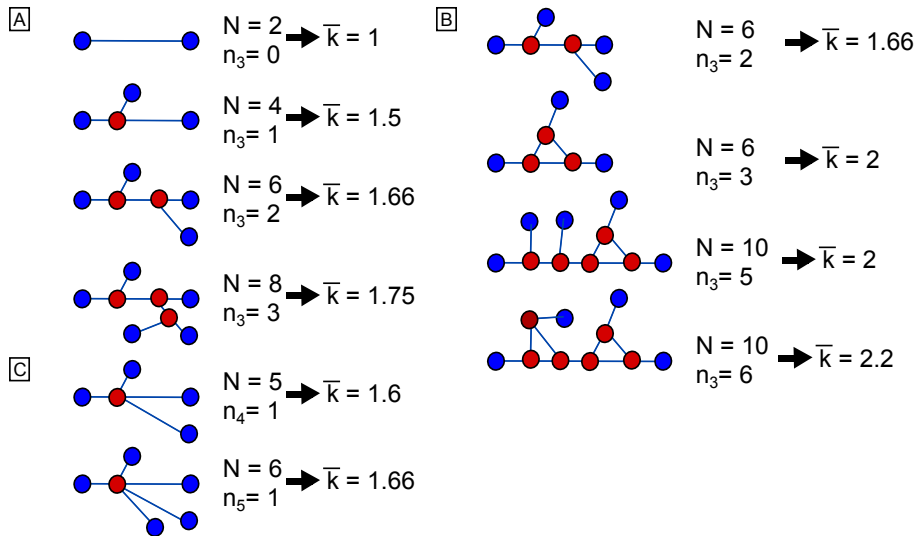
It induces a direct relation between the average vertex degree  $\bar{k}$  and  $N$ :

$$\bar{k} = \frac{2(N-1)}{N} \quad (12)$$

which involves that  $\bar{k} \rightarrow 2$  when  $N \rightarrow \infty$ . Moreover, if we authorize vertex degrees  $> 3$ , it does not change this observation as each edge addition involves also the addition of a new vertex of degree 1 (Fig. 14C). Thus, a value  $\bar{k} \geq 2$  indicates a network that shows cycles. Also, as soon as one cycle is introduced in the network,  $\bar{k}$  is only increasing if new cycles or vertices with degree  $> 3$  are introduced (Fig. 14).

Like for lengths, the coefficient of variation of vertex degrees  $CV_k$  can be computed to characterize the dispersion of degrees. It is expressed in percentage and corresponds to the ratio of the standard deviation  $\sigma_k$  to the mean  $\bar{k}$ :

$$CV_k = \frac{\sigma_k}{\bar{k}} * 100 \quad (13)$$



**Figure 14** Illustration of the average vertex degree meaning for reduced graphs. (A) and (C) For tree graphs,  $\bar{k}$  is increasing toward 2 as the size of the tree graph increases, but the value 2 is never reached. (B) As soon as one cycle is introduced in the network,  $\bar{k} = 2$ . Then,  $\bar{k}$  is only increasing with the addition of new cycles or of nodes of degree  $> 3$ .

The computed values of  $CV_k$  are relatively high and range from 35% to 60%. Indeed, vertex degrees globally range between 1 and 3, sometimes (but rarely) up to 5. As we work with reduced graphs, all vertices of degree  $k_i = 2$  have been removed. Computing an entropy of vertex degrees in this context is not really relevant.

### Correlation of vertex degrees: assortativity

The correlation of degrees between first neighbour vertices has been found to play an important role in structural and dynamical network properties [Maslov and Sneppen, 2002]. In order to assess the correlation of degree of neighbour vertices in our reduced networks, we compute the Pearson correlation coefficient of the degrees at both ends of an edge [Newman, 2002]:

$$r_k = \frac{\frac{1}{S} \sum_{j>i} k_i k_j w_{ij} - [\frac{1}{S} \sum_{j>i} \frac{1}{2}(k_i + k_j) w_{ij}]^2}{\frac{1}{S} \sum_{j>i} \frac{1}{2}(k_i^2 + k_j^2) w_{ij} - [\frac{1}{S} \sum_{j>i} \frac{1}{2}(k_i + k_j) w_{ij}]^2} \quad (14)$$

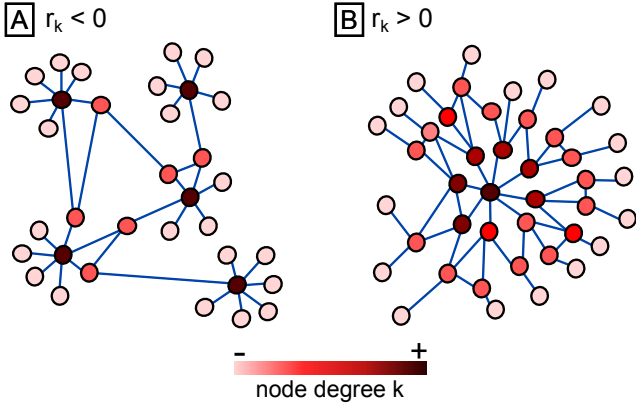
where  $S$  is the total number of edges, and  $w_{ij}$  refers to the corresponding values in the adjacency matrix representation. If  $r < 0$  the network is disassortative, i.e., vertices of high degree tend to connect to vertices of low degrees (Fig. 16A). If  $r = 0$  there is no correlation between vertex degrees. If  $r > 0$  the network is assortative, i.e., vertices of high degrees tend to connect with vertices of high degrees (Fig. 16B).

The values obtained on the 34 reduced networks are all positive, ranging from 0.44 to 0.88. The reduced representations of karstic system are thus assortative (Table 5). This is probably linked to the fact that maximal node degrees rarely exceed 4. Thus, nodes are globally regrouped with nodes of similar low degrees.

### Average shortest path length

In the reduced networks, we define a path as a sequence of edges linking two seed vertices  $i$  and  $j$ . Because all the edges have a length equal to 1 in the reduced graph, the length of a particular path is given by the number of edges that constitute





**Figure 16** Correlation of vertex degrees is a measure of assortativity. (A) An example of disassortative network ( $r_k < 0$ ): vertices of high degree tend to connect to vertices of low degrees; (B) an example of assortative network ( $r_k > 0$ ): vertices of high degree tend to connect to vertices of high degrees ('hubs connected to hubs').

the path. The shortest path between vertices  $i$  and  $j$  is therefore the path that links the two vertices with the minimal number of edges. If no path exists between two vertices, i.e., the network is composed by disconnected components, the length is set to  $\infty$ . Therefore, in order to avoid divergence of the coefficient, we compute the shortest path for each connected components of the network separately and then compute the mean of all components. For any vertex  $i$ , the shortest path length  $SPL_i$  is the average shortest path that separates  $i$  from any other vertex  $j$  of the same connected component:

$$SPL_i = \frac{1}{(N-1)} \sum_j L_{ij} \quad (15)$$

where  $N$  is the number of vertices in the network. We define the average shortest path length of a network as the mean of the shortest path length ratios:

$$\overline{SPL} = \frac{1}{N} \sum_i SPL_i \quad (16)$$

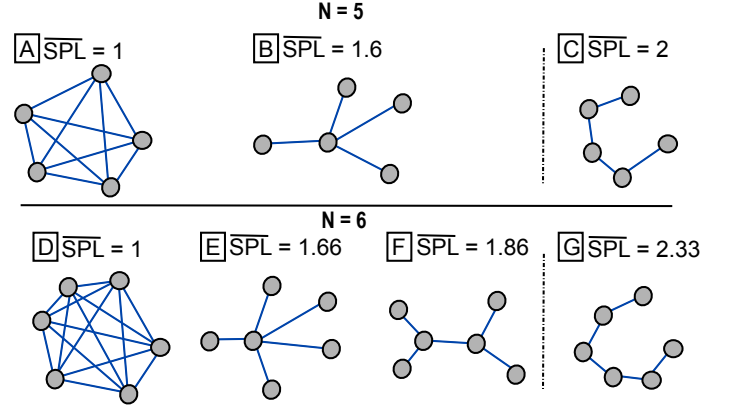
The average shortest path length  $\overline{SPL}$  ranges from 2.32 to 55.76 for the 34 studied networks with a mean value of 16.93. The  $\overline{SPL}$  is related to the efficiency of transfer processes through the network. For example, applied on the world wide web, a short  $\overline{SPL}$  accelerates the transfer of information. Here, it is calculated on the reduced representations of the karst networks, which ignores real distances. It just characterizes the efficiency of the network in terms of its structure or topology. The average shortest path length is thus jointly influenced by two parameters (i) the linearity of the network, and (ii) the size of the network (Figs. 17 and 18).

### Betweenness centrality and central point dominance (CPD)

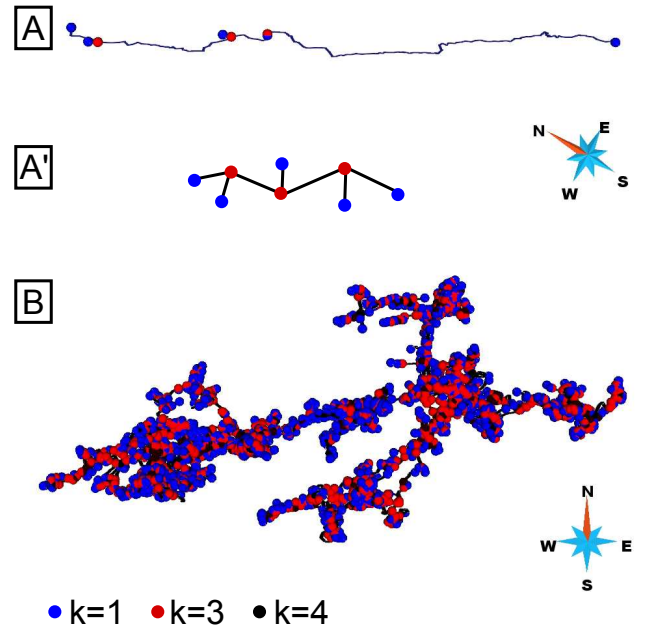
The betweenness centrality of a vertex  $i$  is a measure of the vertex importance in a network. We consider it as a topological measure and therefore compute it on the reduced network. The betweenness centrality relates to the number of shortest paths that cross a vertex  $i$  with the following relation:

$$B_i = \sum_{jk} \frac{\xi_{j,k}(i)}{\xi_{j,k}} \quad (17)$$

where  $\xi(j, i, k)$  is the number of shortest paths between the vertices  $j$  and  $k$  that pass through the vertex  $i$ , and  $\xi(j, k)$  is the total number of shortest paths between  $j$  and  $k$ . The sum takes over all distinct pairs  $j, k$  of vertices. The  $B_i$  indexes the potential of a point for control. As it is essentially a count, its magnitude depends, among others, upon the number of points in the graph. To eliminate this impact, Freeman [1977] proposed to use a relative centrality,  $B_i^*$ . For any undirected star graph, this normalization has to guarantee that the relative



**Figure 17** Average shortest path length,  $\overline{SPL}$ : (A) and (C) are complete graphs, each node is directly linked to all other ones by one edge, thus  $\overline{SPL}_A = \overline{SPL}_C = 1$  independently of the total number of nodes  $N$ . On the opposite, (C) and (G) are linear graphs, but with different sizes: (C) is longer than (G), thus  $\overline{SPL}_B < \overline{SPL}_D$ . As we compute  $\overline{SPL}$  on a reduced graph, these linear structures cannot be observed (all nodes of degree 2 have been removed). Cases (B) and (E) demonstrate that, for the same spatial architecture,  $\overline{SPL}$  increases with the total number of nodes  $N$ . Cases (E) and (F) demonstrate that, for the same total number of nodes  $N$ ,  $\overline{SPL}$  also varies depending on the architecture of the tree graph: case (F) is more linear than case (E).



**Figure 18** Examples of karstic networks and their associated  $\overline{SPL}$  value. (A) and (A') The Clydach Gorge Caf1 network is a short one with  $\overline{SPL} = 2.32$ ; (A) presents the real network with only the seed vertices visible. (A') presents the nonscale corresponding view of its reduced representation. (B) The Lechuguilla cave has  $\overline{SPL} = 55.76$ .

betweenness centrality of the central point is equal to 1. Thus:

$$B_i^* = \frac{B_i}{N^2 - 3N + 2} \quad (18)$$

where  $N$  is the number of nodes in the network. The central point dominance for a network is then defined by Freeman [1977] as

$$CPD = \frac{1}{N-1} \sum_i (B_{max}^* - B_i^*) \quad (19)$$

where  $B_{max}^*$  is the largest value of relative betweenness centrality in the network.

The central point dominance is 0 for a complete graph, i.e., a graph in which each vertex has a direct edge to all the other vertices, and 1 for a star graph in which a central vertex is included in all paths (Figs. 19A and B). On karstic systems, we obtained values ranging from 0.02 to 0.57 with a mean value of 0.39. The highest value is obtained on the Eglwys Faen cave, which is characterized by several vertices of degree  $k \geq 4$  and quite centralized. On the opposite, the Mammuthöhle cave, which is composed by 25 connected components, all of them without noticeable centralized organization, get the lowest value of our data set with  $CPD = 0.02$  (Fig. 20).

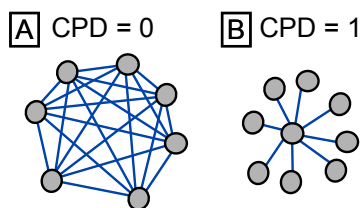
## COMPARISON OF THE METRICS AND DISCUSSION

We proposed different metrics for the characterization of karst geometry and topology. All of them have been computed on 34 karstic networks coming from various parts of the world and related to different speleogenetic processes. In this section, we study the relation between all these metrics in order to identify redundant ones and to better understand their signification. To start the analysis, Table 6 presents the linear correlation coefficients that we computed between the different metrics.

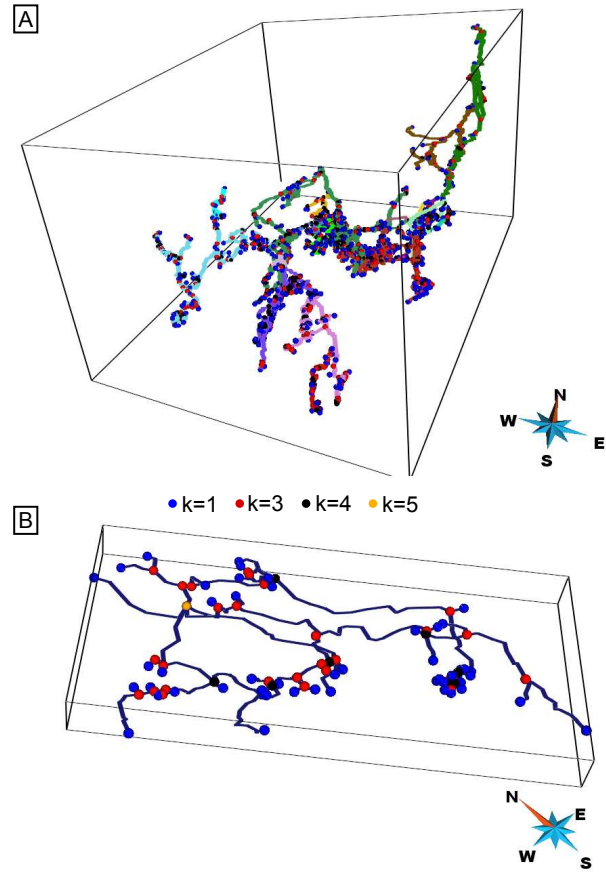
### Geometrical parameters

We chose to focus the geometrical analysis on orientation, length, and tortuosity of conduits.

The orientation analysis helps in identifying preferential direction of speleogenesis that can be linked to the existence of inception features, like fractures, bedding, and faults. Orientation also underlines the curvilinear patterns of karsts developed along almost horizontal bedding planes. Our results show the high variety of orientation patterns and the difficulty to define general reference cases. The use of orientation entropy, as recently proposed for urban street analysis [Gudmundsson and Mohajeri, 2013], provides an interesting tool to quantitatively classify network on orientation criteria. However, density map analysis is necessary to complement this metric and especially



**Figure 19** Central point dominance (CPD): (A) example of a complete graph, i.e., a graph in which each vertex has a direct edge to all the other vertices:  $CPD = 0$ ; (B) for a star graph  $CPD = 1$ , i.e., a central vertex is included in all paths.



**Figure 20** Examples of central point dominance (CPD) results: (A) Mammuthöhle cave is constituted by 25 connected components and a dispersed organisation  $CPD = 0.02$ ; (B) Eglwys Faen is characterized by the existence of several high degree nodes ( $k_{max} = 5$ ) and a more centralized pattern  $CPD = 0.57$ .

to detect the existence of a vertical preferential development of conduits.

The average length, length entropy, and the coefficient of variation of conduit length are interesting parameters to describe the geometry of karstic networks. The length entropy informs about the distribution of conduit dimensions, while the coefficient of variation shows the extent of variability relative to the mean. A negative nonlinear correlation is observed between the length entropy and the size of the network (expressed through the numbers of survey stations  $n$ , of lines of sight  $s$ , of seed vertices  $N$ , and of karst branches  $S$ ). The computation of entropy divides indeed the samples in bins. Thus, for smallest karsts, the probability to obtain a bin that contains one or two samples is higher. So, even if the population is tightened around the mean (low value of the coefficient of variation), the resulting distribution is more likely uniform.

Tortuosity, also called sinuosity index, has often been used to describe karst geometry [Pardo-Iguzquiza et al., 2011, Jeannin et al., 2007]. Nonetheless, the computed values reveal only slight variations that do not seem to correspond to a particular or clear curvilinear karst pattern classification when visualizing the corresponding systems. We suspect that tortuosity is probably strongly affected by the survey methodology. If this is true, it should be used as a criterion or as an input parameter for conduit geometry simulation only when data were acquired following a very well-defined sampling procedure, i.e., with stations located in the center of the conduits and a fixed distance between the stations.

**Table 6**

Linear (Pearson) correlation coefficients  $r$  between all studied metrics; light blue values correspond to  $0.7 \leq |r| < 0.85$ ; dark blue values correspond to  $|r| \geq 0.85$ , which expresses a strong linear correlation between both variables

	$n$	$s$	$H_0$	$\frac{1}{\bar{l}_{en}}$	$H_{len}$	$CV_{len}$	$\bar{t}$	$N$	$S$	$p$	$N_{cycl}$	$\alpha$	$\beta$	$\gamma$	$D_C$	$\theta$	$\bar{k}$	$CV_k$	$r_k$	$\overline{SPL}$	$CPD$
$n$	1.00	0.32	-0.03	-0.49	-0.01	0.20	0.87	0.85	0.27	0.69	0.13	0.26	0.01	0.15	0.08	0.26	-0.03	0.37	0.88	-0.04	
$s$	1.00	0.32	-0.04	-0.49	-0.02	0.20	0.88	0.86	0.27	0.71	0.14	0.27	0.03	0.16	0.09	0.27	-0.03	0.38	0.88	-0.04	
$H_0$	0.32	0.32	-0.48	-0.15	-0.04	0.24	0.28	0.28	0.15	0.25	0.38	0.57	-0.04	0.34	0.36	0.57	-0.42	0.58	0.38	-0.39	
$\frac{1}{\bar{l}_{en}}$	-0.03	-0.04	-0.48	0.12	0.05	-0.08	-0.09	-0.09	-0.15	-0.11	-0.32	-0.41	0.00	-0.27	-0.30	-0.41	0.18	-0.42	-0.05	0.32	
$H_{len}$	-0.49	-0.49	-0.15	0.12	-0.69	-0.24	-0.36	-0.34	-0.29	-0.24	-0.11	-0.24	-0.05	-0.15	-0.10	-0.24	0.01	-0.23	-0.38	0.05	
$CV_{len}$	-0.01	-0.02	-0.04	0.05	-0.69	0.04	-0.13	-0.14	-0.10	-0.18	-0.05	0.02	-0.01	-0.01	0.00	0.02	-0.03	-0.14	0.07	0.23	
$\bar{t}$	0.20	0.20	0.24	-0.08	-0.24	0.04	0.11	0.10	0.17	0.08	0.19	0.25	0.10	0.20	0.16	0.25	-0.07	0.17	0.14	0.07	
$N$	0.87	0.88	0.28	-0.09	-0.36	-0.13	0.11	1.00	0.32	0.95	0.27	0.34	0.17	0.28	0.20	0.34	0.02	0.39	0.76	-0.07	
$S$	0.85	0.86	0.28	-0.09	-0.34	-0.14	0.10	1.00	0.32	0.97	0.29	0.36	0.20	0.30	0.22	0.36	0.01	0.39	0.74	-0.08	
$p$	0.27	0.27	0.15	-0.15	-0.29	-0.10	0.17	0.32	0.32	0.35	0.20	0.16	0.07	0.15	0.18	0.16	0.11	0.03	-0.08	-0.67	
$N_{cycl}$	0.69	0.71	0.25	-0.11	-0.24	-0.18	0.08	0.95	0.97	0.35	0.39	0.42	0.33	0.40	0.33	0.42	-0.03	0.37	0.59	-0.11	
$\alpha$	0.13	0.14	0.38	-0.32	-0.11	-0.05	0.19	0.27	0.29	0.20	0.39	0.94	0.86	0.99	0.99	0.94	-0.76	0.65	0.06	-0.34	
$\beta$	0.26	0.27	0.57	-0.41	-0.24	0.02	0.25	0.34	0.36	0.16	0.42	0.94	0.72	0.95	0.91	1.00	-0.77	0.79	0.26	-0.25	
$\gamma$	0.01	0.03	-0.04	0.00	-0.05	-0.01	0.10	0.17	0.20	0.07	0.33	0.86	0.72	0.90	0.85	0.72	-0.63	0.38	-0.07	-0.07	
$D_C$	0.15	0.16	0.34	-0.27	-0.15	-0.01	0.20	0.28	0.30	0.15	0.40	0.99	0.95	0.90	0.97	0.95	-0.76	0.66	0.11	-0.24	
$\theta$	0.08	0.09	0.36	-0.30	-0.10	0.00	0.16	0.20	0.22	0.18	0.33	0.99	0.91	0.85	0.97	0.91	-0.82	0.60	0.02	-0.35	
$\bar{k}$	0.26	0.27	0.57	-0.41	-0.24	0.02	0.25	0.34	0.36	0.16	0.42	0.94	1.00	0.72	0.95	0.91	-0.77	0.79	0.26	-0.25	
$CV_k$	-0.03	-0.03	-0.42	0.18	0.01	-0.03	-0.07	0.02	0.01	0.11	-0.03	-0.76	-0.77	-0.63	-0.76	-0.82	-0.77	-0.63	-0.09	0.21	
$r_k$	0.37	0.38	0.58	-0.42	-0.23	-0.14	0.17	0.39	0.39	0.03	0.37	0.65	0.79	0.38	0.66	0.60	0.79	-0.63	0.43	-0.21	
$\overline{SPL}$	0.88	0.88	0.38	-0.05	-0.38	0.07	0.14	0.76	0.74	-0.08	0.59	0.06	0.26	-0.07	0.11	0.02	0.26	-0.09	0.43	0.23	
$CPD$	-0.04	-0.04	-0.39	0.32	0.05	0.23	0.07	-0.07	-0.08	-0.67	-0.11	-0.34	-0.25	-0.07	-0.24	-0.35	-0.25	0.21	-0.21	0.23	

Other geometrical metrics have been investigated in previous papers [Howard, 1971, Pardo-Iguzquiza et al., 2011]. In particular, a set of metrics use widths, passage areas, mean height diameters. These metrics provide important volumetric information, but they required field data that were not available for all the karst systems considered in this study and this is why they are not included here. Such a volumetric description is, however, clearly vital for providing a complete description of a cave system.

## Topological parameters

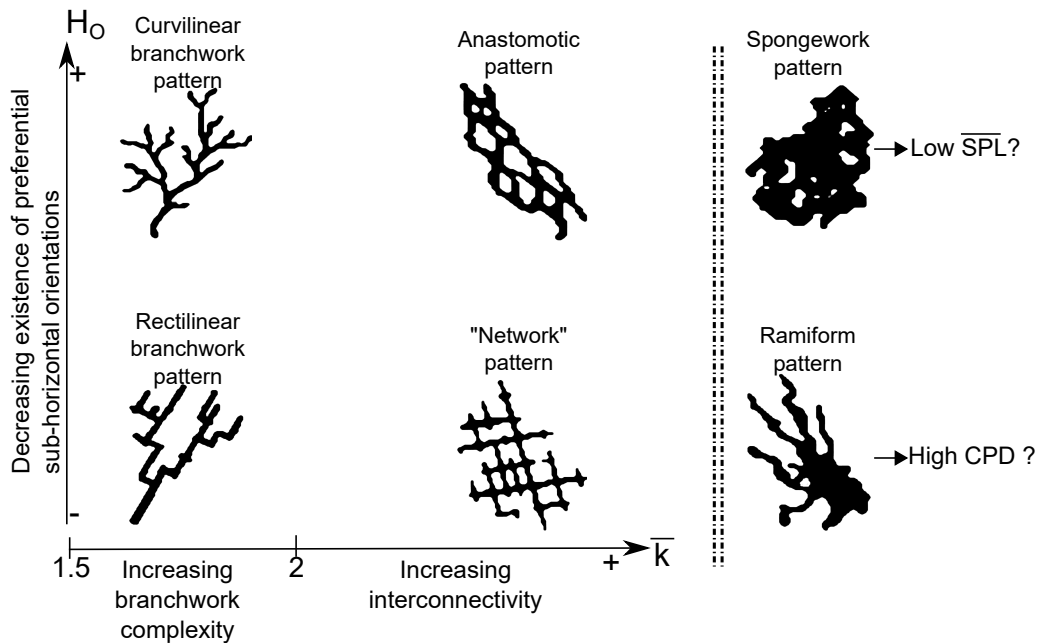
Concerning topology, we used several metrics to describe the network organization. Considering their correlation coefficients, some of them appear to express similar characteristics of the networks. It is important to remember that topological metrics are computed on reduced graphs: all vertices of degree  $k = 2$  are removed from the networks.

The interconnectivity of the network is jointly expressed by all Howard's parameters  $\alpha$ ,  $\beta$ , and  $\gamma$ , the resulting degree of connectivity  $D_C$ , the global cyclic coefficient  $\theta$ , and finally, the average vertex degree  $\bar{k}$ . All these parameters are indeed highly linearly correlated (Table 6). Additionally, a perfect linear correlation ( $r = 1$ ) is observed between  $\beta$  and  $\bar{k}$ , which are linked by the relation  $\bar{k} = 2\beta$ . As a consequence, we would recommend to keep only one of these metrics for characterising the interconnectivity of the network. Our choice goes to the average vertex degree  $\bar{k}$ , which is quite easy to compute. As demonstrated in section 4.2.1,  $\bar{k}$  has the advantage to classify tree graphs (with values  $1.5 \leq \bar{k} < 2$ ) and interconnected systems (values  $\bar{k} \geq 2$ ). Note that these peculiarities of  $\bar{k}$  are completely linked to the fact that we compute it on reduced graphs. In other studies, e.g., on fracture networks [Andresen

et al., 2013], the presence of nodes of degree 2 leads to values  $\bar{k} \leq 1.5$ .

The correlation of vertex degrees have demonstrated that the reduced graphs of karstic systems were all assortative ( $0.44 \leq r_k \leq 0.88$ ). Previously, Newman [2002] has showed that many social networks were assortative while technological and biological networks seem to be disassortative. It is perhaps linked to the fact that high degree vertices are quite rare in karstic networks: we did not record  $k$  values  $> 5$ . As a comparison, the equivalent graphs of transformed fracture networks proposed by Andresen et al. [2013] were characterized by a mean value  $k_{max} = 29$ . Assortativity is slightly linked to  $\bar{k}$  ( $r = 0.79$ , Table 6), but the low correlation coefficients obtained with other metrics relating to interconnectivity ( $\alpha$ ,  $\gamma$ ,  $D_C$ ,  $\theta$ ) show that it does not express the interconnectivity of the network. It is a complementary metric.

Average shortest path length ( $\overline{SPL}$ ) is a standard metric in graph theory [e.g., Chandy and Misra, 1982, Watts and Strogatz, 1998, Albert and Barabási, 2002, Newman, 2003, Andresen et al., 2013]. Expressing a kind of transfer efficiency through a network, the average shortest path length is jointly influenced by the size of the network and its architecture: it varies as the network is spread (almost linear,  $\overline{SPL} > 1$ ) or hunched ( $\overline{SPL} = 1$ ). This explains the high correlation coefficients we obtained with the network size parameters  $n$ ,  $s$ ,  $N$ , and  $S$  (Table 6). We computed it on the reduced graphs where edge lengths equal 1. But alternative definitions could be developed and used. Firstly, edge length could be defined as the curvilinear length of the corresponding branch length to more accurately express the network efficiency, in terms of flow. This would probably reinforce the correlation between network size and the new metric. Secondly, considering the cave entrance(s) or outlet(s) and a downgradient could be in-



**Figure 21** Common patterns of solutional caves according to Palmer [1991] and links with some topological metrics we propose: interconnectivity is increasing with the average node degree  $\bar{k}$ ; orientation entropy  $H_O$  is lower when clear preferential directions of karst development exist; the central point dominance  $CPD$  should be higher for ramiform patterns, and spongework patterns could be characterized by low values of average shortest path length  $\overline{SPL}$ . Nonetheless, especially for the two last parameters, relations are still required to be demonstrated with a dedicated statistical analysis supported by speleogenetical considerations.

teresting to get a metric with a more hydrogeological meaning. In practice, siphons or other kinds of conduits could add complexity to automatically define a downgradient. In the present study we do not always get the entrance(s)/outlet(s) for the studied networks, but it could be an interesting perspective.

Central point dominance allows us to classify karstic networks on a new topological characteristic: theoretically restricted between 0 and 1, values obtained for karstic systems range from 0.02 to 0.57; no noticeable correlation is observed with the other metrics showing nonredundancy of the information provided by the  $CPD$ , which expresses the potential centralized architecture of the network.

### About karst patterns

Trying to relate the proposed topological metrics to standard patterns of karstic systems as defined by Palmer [1991] and/or to a geological environment is tempting (Fig. 21). Pure branchwork patterns must have  $1.5 \leq \bar{k} < 2$ , with values increasing with the number of branches. Maze patterns, gathering anastomotic and network patterns of Palmer's classification, should be characterized by a high interconnectivity (and thus  $\bar{k} > 2$ ) as well as spongework and, perhaps, ramiform patterns. Ramiform patterns should probably distinguish themselves by a higher value of  $CPD$  and, potentially, a high assortativity. Spongework patterns should probably show a low value of  $\overline{SPL}$  as they seem quite hunched. But all those intuitive interpretations cannot be rigorously demonstrated here. Indeed, it would require us to accurately attach each system to one pattern. To do so, a visual analysis of the karst network alone is not sufficient and necessitates a parallel field study. Moreover, most of the studied karst systems appear to be the result of different speleogenetic phases. As a consequence, relating parameters to a proper geologic environment would require us to split the networks in homogeneous speleogenetic subparts,

which is again questionable without a field study. Splitting networks in various groups reduces the population of each group, and the statistical study should so ensure getting a minimum amount of networks in each group. Such work constitutes an interesting perspective that could be addressed with the help of speleologists.

### About sampling errors

Numerical tools have been developed to automate the measures on graph representations of the networks. But cave surveys and data acquisition processes still induce errors that cannot be completely automatically corrected. It concerns errors coming from nonmeasured conduits: potentially relating to small diameter conduits it should probably affect metrics on volumetric characteristics but not necessarily those concerning the geometry. It also concerns errors coming from the fact that data comes from an exploration realized with a different aim than a statistical one: location of the survey station inside the conduit, sampling step, sampling strategy for large caves. Different parts of the Sieben Hengste network have been studied. The SP1 and SP2 are two subparts of the Sieben Hengste LargePart network. It is interesting to notice that, despite the fact that the large part network is obviously segmented in different zones (Fig. 1), most of the values computed for both subparts are close to those computed for the large network. The main differences are indeed observed for metrics that are sensitive to the size of the complete or reduced graph, such as  $\bar{k}$ ,  $H_{len}$  (and obviously  $n$ ,  $s$ ,  $N$  and  $S$ ). Even if these results could be interpreted as an encouraging point, the impact of survey errors remains to be further explored. One way to address it could be to perform repeated random removals of network parts in a Monte Carlo framework. Then, computing the different metrics on the resulting subsystems would allow us to assess the sensitivity of the metrics to incomplete data sets.

## CONCLUSION

We have proposed and computed 21 metrics on a data set of more than 30 karstic networks. The data set includes some of the largest explored cave systems in the world and represents a broad range of geological and speleogenetic conditions allowing us to test the proposed metrics, their variability, and their usefulness for the discrimination of different morphologies.

Numerical tools have been developed to automate these computations on two graph representations of the networks, consisting of vertices and edges (i) a complete version of the network for geometrical metrics, and (ii) a reduced version of the network for topological considerations. These tools are available on demand. Among the tested metrics, eight appear particularly relevant to describe the karstic networks. For geometry, orientation entropy ( $H_O$ ), completed by a density map analysis, provides a useful tool to quantitatively classify networks on orientation criteria and to help detect the existence of inception features. It is interestingly completed by average lengths ( $\overline{len}$ ), coefficient of variations of lengths ( $CV_{len}$ ), and length entropy of conduits ( $H_{len}$ ), which describe the extension of the conduits within the network. Concerning topology, we would recommend computing the average vertex degree of the reduced graph of networks ( $\overline{k}$ ) to characterize the interconnectivity of the systems. It is indeed simple to compute, it discriminates properly the interconnected systems (mazes) from the acyclic ones (tree-like structures), and it permits us to classify the acyclic systems as a function of the total number of branches. Assortativity ( $r_k$ ), average shortest path length ( $SPL$ ), and central point dominance ( $CPD$ ) are metrics inspired by graph theory that have been used for the first time to characterize karsts. These parameters express various complementary aspects of network topologies, and we gave some examples to help the user in interpreting their physical meaning.

Besides the metrics, the main contribution of this paper is to provide a reference database of values measured on several actual karstic networks. Such a toolbox and associated database offer the possibility to test karst network models on a quantitative level when trying to qualify their degree of 'realism'. In that perspective, feeding this database should be encouraged to increase its representativeness.

## ACKNOWLEDGEMENTS

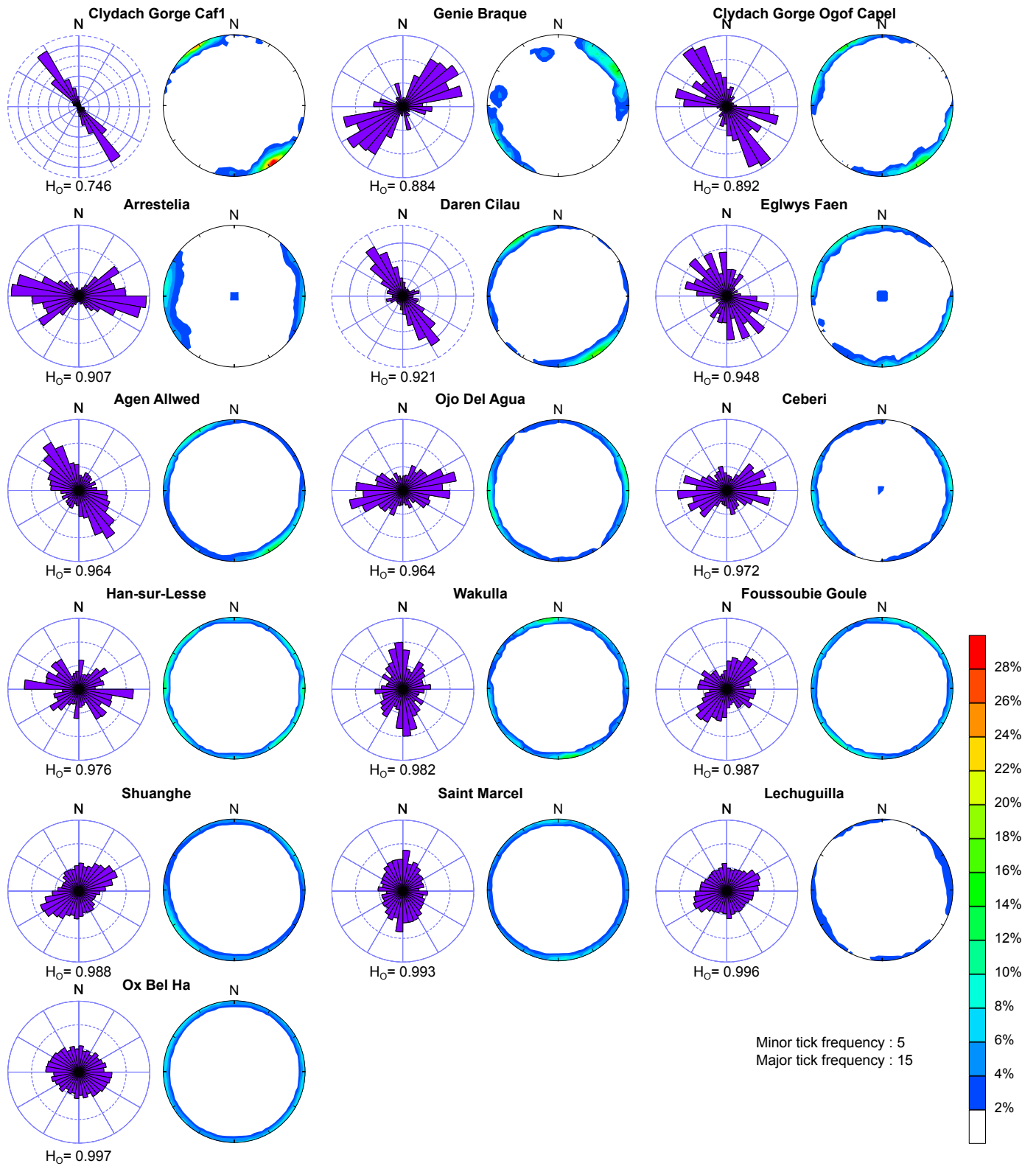
A large part of this work was performed in the frame of the RING project at Université de Lorraine. We would like to thank the industrial and academic sponsors of the RING-Gocad Research Consortium managed by ASGA for their support. We also thank Paradigm for providing the SKUA-GOCAD software and API. We also would like to thank all the colleagues that have provided us with 3D karst data sets: J. Bodin, J. Botazzi, J-P. Cassou, S. Jailliet, P-Y. Jeannin, T. Kincaid, E. Pardo-Iguzquiza, and Quintana Roo Speleological Survey. We also thank S. Viseur for interesting discussions shared about this work during various conferences. Richard Marston, co-editor of Geomorphology, L. Piccini and other anonymous reviewers are also thanked for their constructive remarks, which helped to improve this manuscript.

## REFERENCES

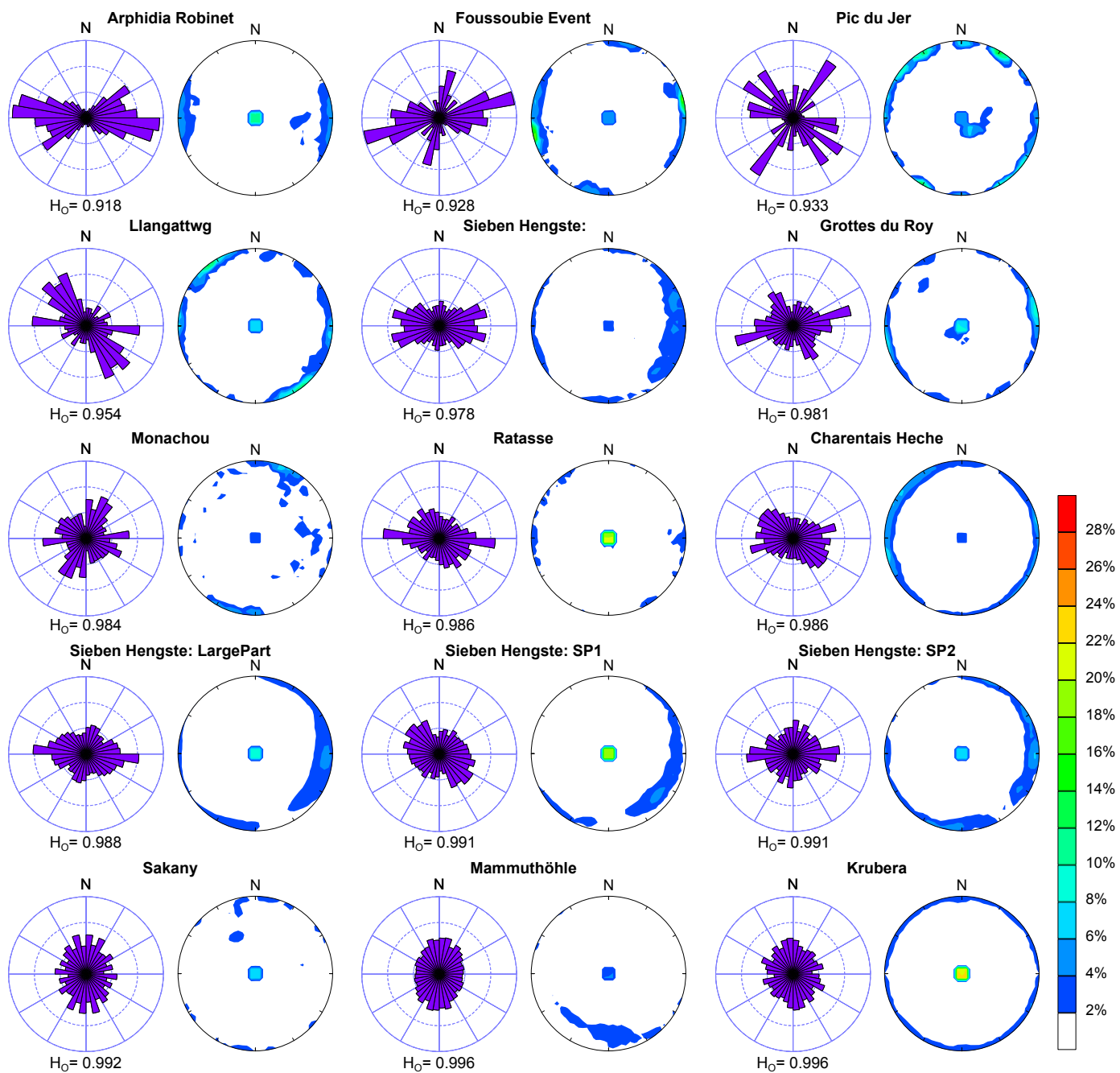
- Albert, R., Barabási, A.-L., jan 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics* 74 (1), 47–97. (Cited page 18)
- Alexander, S. C., Larson, E., Greenwaldt, B., Alexander, E. C., Rahimi, M., Hall, M., Ave, B., Paul, S., 2013. Combining Lidar, Aerial Photography, and Pictometry ® Tools for Karst Features Database Management. In: 13th sinkhole conference. pp. 441–448. (Cited page 1)
- Andresen, C. A., Hansen, A., Le Goc, R., Davy, P., Hope, S. M., 2013. Topology of fracture networks. *Frontiers in Physics* 1, 1–5. (Cited pages 12 and 18)
- Beekman, F., Badsì, M., van Wees, J.-D., may 2000. Faulting, fracturing and in situ stress prediction in the Ahnet Basin, Algeria — a finite element approach. *Tectonophysics* 320 (3-4), 311–329. (Cited page 6)
- Billaux, D., Chiles, J., Hestir, K., Long, J., jul 1989. Three-dimensional statistical modelling of a fractured rock mass—an example from the Fanay-Augères mine. *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts* 26 (3-4), 281–299. (Cited page 6)
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D., feb 2006. Complex networks: Structure and dynamics. *Physics Reports* 424 (4-5), 175–308. (Cited page 2)
- Bonniver, I., 2011. Etude hydrogéologique et dimensionnement par modélisation du "Système-Traçage" du réseau karstique de Han-sur-Lesse (Massif de Boine - Belgique). Ph.D. thesis, FUNDP Namur. (Cited page 8)
- Borghi, A., Renard, P., Jenni, S., 2012. A pseudo-genetic stochastic model to generate karstic networks. *Journal of Hydrology* 414-415, 516 – 529. (Cited page 1)
- Caverne, J.-B., 2011. U-GPS: un prolongement du système GPS sous terre. *XYZ* 126, 27–36. (Cited page 2)
- Chandy, K. M., Misra, J., nov 1982. Distributed computation on graphs: shortest path algorithms. *Communications of the ACM* 25 (11), 833–837. (Cited page 18)
- Chen, Z., Goldscheider, N., jun 2014. Modeling spatially and temporally varied hydraulic behavior of a folded karst system with dominant conduit drainage at catchment scale, Hochifèn–Gottesacker, Alps. *Journal of Hydrology* 514, 41–52. (Cited page 1)
- Collon-Drouaillet, P., Henrion, V., Pellerin, J., nov 2012. An algorithm for 3D simulation of branchwork karst networks using Horton parameters and A\* Application to a synthetic case. *Geological Society, London, Special Publications* 370 (1), 295–306. (Cited page 1)
- Costa, L. F. d. F., Rodrigues, F. A., Travieso, G., Villas Boas, P. R., 2007. Characterization of complex networks: A survey of measurements. *Advances in Physics* 56 (1), 167–242. (Cited page 2)
- Curl, R. L., 1966. Caves as a measure of karst. *The Journal of Geology* 74 (5), 798–830. (Cited page 1)
- Filipponi, M., Jeannin, P. P.-Y., Tacher, L., may 2009. Evidence of inception horizons in karst conduit networks. *Geomorphology* 106 (1-2), 86–99. (Cited page 8)
- Ford, D., Williams, P., 2007. *Karst Hydrogeology and Geomorphology*. John Wiley and Sons, Ltd. (Cited page 7)
- Fournillon, A., Abelard, S., Viseur, S., Arfib, B., Borgomano, J., 2012. Characterization of karstic networks by automatic extraction of geometrical and topological parameters: comparison between observations and stochastic simulations. *Geological Society, London, Special Publications* 370 (1996), 247–264. (Cited page 2)
- Freeman, L. C., feb 1977. A set of measures of centrality based on betweenness. *Sociometry* 40 (1), 35–41. (Cited pages 10, 16, and 17)
- Garrison, W. L., Marble, D. F., 1962. The structure of transportation networks. Tech. rep., US Army Transport Command Technical Report. (Cited page 10)
- Glennon, A., Groves, C., 2002. An examination of perennial stream drainage patterns within the Mammoth Cave Watershed, Kentucky. *Journal of Cave and Karst Studies* 64 (1), 82–91. (Cited page 5)
- Glennon, J. A., 2001. Application of morphometric relationships to active flow networks within the Mammoth Cave watershed. Tech. Rep. November, Western Kentucky University. (Cited page 5)
- Gudmundsson, A., Mohajeri, N., nov 2013. Entropy and order in urban street networks. *Scientific Reports* 3. (Cited pages 7 and 17)
- Hendrick, M., Renard, P., nov 2016. Subnetworks of Percolation Backbones to Model Karst Systems Around Tulum, Mexico. *Frontiers in Physics* 4. (Cited page 1)
- Horton, R. E., 1945. Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Bulletin Geological Society of America* 56, 275–370. (Cited page 1)
- Howard, A. D., 1971. Quantitative Measures of Caves Patterns. *Caves and Karst* 13 (1), 1–7. (Cited pages 1, 5, 10, 12, and 18)
- Howard, A. D., Keetch, M. E., Vincent, C. L., 1970. Topological and geometrical properties of braided patterns. *Water Resources Research* 6 (6), 1674–1688. (Cited pages 1, 10, and 12)

- Jaillet, S., Sadier, B., Arnaud, J., Azma, M., Boche, E., Cailhol, D., Filipponi, M., Roux, P., Varrel, E., 2011. Topographie, représentation et analyse morphologique 3D de drains, de conduits et de parois du karst. In: Jaillet, S., Ployon, E., Villemin, T. (Eds.), *Images et modèles 3D en milieux naturels*, edytem Edition. CNRS, pp. 119–130. (Cited page 2)
- Jaquet, O., Siegel, P., Klubertanz, G., Benabderrhamane, H., 2004. Stochastic discrete model of karstic networks. *Advances in Water Resources* 27 (7), 751–760. (Cited page 1)
- Jeannin, P.-Y., 1996. Structure et comportement hydraulique des aquifères karstiques. Ph.D. thesis, Université de Neuchâtel. (Cited page 8)
- Jeannin, P.-Y., feb 2001. Modeling flow in phreatic and epiphreatic Karst conduits in the Hölloch Cave (Muotatal, Switzerland). *Water Resources Research* 37 (2), 191–200. (Cited page 1)
- Jeannin, P. Y., Groves, C., Häuselmann, P., 2007. Chapter 3: Speleological investigations. In: Goldscheider, N., Drew, D. (Eds.), *Methods in Karst Hydrogeology*. Taylor & Francis, London. (Cited pages 2, 10, and 17)
- Journel, A. G., Deutsch, C. V., apr 1993. Entropy and spatial disorder. *Mathematical Geology* 25 (3), 329–355. (Cited page 7)
- Kansky, K. J., 1963. Structure of transportation networks: relationships between network geometry and regional characteristics. Tech. rep., University of Chicago, Chicago, US. (Cited page 10)
- Kim, H.-J., Kim, J. M., sep 2005. Cyclic Topology in Complex Networks. *Physical review. E, Statistical, nonlinear, and soft matter physics* 72 (3 Pt 2), 2–5. (Cited pages 10 and 12)
- Király, L., Mathey, B., Tripet, J.-P., 1971. Fissuration et orientation des cavités souterraines - Région de la grotte de Milandre (Jura tabulaire). *Bulletin de la société neuchateloise des sciences naturelles* 94 (3), 100–114. (Cited page 6)
- Labourdette, R., Lascu, I., Mylroie, J., Roth, M., 2007. Process-like modeling of flank-magin caves: from genesis to burial evolution. *Journal of Sedimentary Research* 77 (10), 965–979. (Cited page 1)
- Maslov, S., Sneppen, K., 2002. Specificity and stability in topology of protein networks. *Science* 296 (5569), 910–913. (Cited page 15)
- Newman, M. E. J., 2002. Assortative mixing in networks. *Physical review letters* 89 (20), 208701. (Cited pages 10, 15, and 18)
- Newman, M. E. J., 2003. The Structure and Function of Complex Networks. *Society for Industrial and Applied Mathematics (SIAM) Review* 45 (2), pp. 167–256. (Cited pages 12 and 18)
- Palmer, A., 1991. Origin and morphology of limestone caves. *Geological Society of America Bulletin* 103, 1–21. (Cited pages 2, 6, 10, and 19)
- Pardo-Iguzquiza, E., Duran-Valsero, J., Rodriguez-Galiano, V., 2011. Morphometric analysis of three-dimensional networks of karst conduits. *Geomorphology* 132 (1-2), 17–28. (Cited pages 2, 5, 7, 10, 17, and 18)
- Pardo-Iguzquiza, E., Dowd, P. A., Chaoshui, X., Duran-Valsero, J. J., Pardo-Iguzquiza, E., Xu, C., Durán-Valsero, J. J., jan 2012. Stochastic simulation of karst conduit networks. *Advances in Water Resources* 35, 141–150. (Cited pages 1 and 10)
- Piccini, L., 2011. Recent Developments on morphometric analysis of karst caves. *Acta Carsologica* 40 (1), 43–52. (Cited pages 2 and 5)
- Ployon, E., Jaillet, S., Barge, O., 2011. Acquisition et traitements de nuages de points 3D, par des techniques légères et à faible coûts, pour l'élaboration de MNT à haute résolution. *Images et modèles 3D en milieux naturels*, 155–168. (Cited page 2)
- Ravasz, E., Barabasi, A.-L., 2002. Hierarchical Organization in Complex Networks. *Physical Review E* 67 (2), 026112. (Cited page 2)
- Rongier, G., Collon-Drouaillet, P., Filipponi, M., may 2014. Simulation of 3D karst conduits with an object-distance based method integrating geological knowledge. *Geomorphology* 217, 152–164. (Cited page 2)
- Sadier, B., 2013. 3D et géomorphologie karstique : La grotte Chauvet et les cavités des Gorges de l'Ardèche. Ph.D. thesis, Université de Grenoble. (Cited page 2)
- Scheidegger, A. E., 1966. Statistical Description of River Networks. *Water Resources Research* 2 (4), 785–790. (Cited page 1)
- Scheidegger, A. E., 1967. On the Topology of River Nets. *Water Resources Research* 3 (1), 103–106. (Cited page 1)
- Schiller, A., Renard, P., 2016. An optical laser device for mapping 3D geometry of underwater karst structures: First tests in the Ox Bel'Ha system, Yucatan, Mexico. *Boletín Geológico y Minero* 127 (1), 99–110. (Cited page 2)
- Smart, J. S., 1969. Topological Properties of Channel Networks. *Geological Society of America Bulletin* 80, 1757–73. (Cited page 1)
- Viseur, S., Jouvès, J., Fournillon, A., Arfib, B., Guglielmi, Y., 2014. 3D stochastic simulation of caves : application to Saint-Sébastien case study ( SE , France ). *Karstologia* 64, 17–24. (Cited page 1)
- Watts, D. J., Strogatz, S. H., jun 1998. Collective dynamics of "small-world" networks. *Nature* 393 (6684), 440–442. (Cited pages 12 and 18)
- Weishampel, J. F., Hightower, J. N., Chase, A. F., Chase, D. Z., Patrick, R. A., 2011. Detection and morphologic analysis of potential below-canopy cave openings in the karst landscape around the Maya polity of Caracol using airborne LiDAR. *Journal of Cave and Karst Studies* 73 (3), 187–196. (Cited page 1)
- Williams, P. W., Williams, 1972. Morphometric Analysis of Polygonal Karst in New Guinea. *Geological Society of America Bulletin*. (Cited page 1)
- Woldenberg, M. J., 1966. Horton's Laws Justified in Terms of Allometric Growth and Steady State in Open Systems. *Bulletin of Geological Society of America* 77, 431–434. (Cited page 1)
- Zhu, J., Taylor, T. P., Currens, James C Crawford, M. M., 2014. Improved karst sinkhole mapping in kentucky using Lidar techniques: a pilot study in Floyds Fork watershed. *Journal of Cave and Karst Studies* 76 (3), 207–216. (Cited page 1)
- Ziman, J. M., 1979. Models of disorder: the theoretical physics of homogeneously disordered systems. CUP Archive. (Cited page 7)
- Zoback, M. L., 1992. First- and second-order patterns of stress in the lithosphere: The World Stress Map Project. *Journal of Geophysical Research* 97 (B8), 11703. (Cited page 6)

## APPENDIX A: COMPLETE RESULTS OF ORIENTATION ANALYSIS



**Figure 22** Orientation analysis results (Part 1): karst networks without a clear vertical preferential direction (i.e., relative frequency of vertical segments is inferior to those of other directions). The karstic systems are classified from the ones with the lower orientation entropy ( $H_o = 0.746$ ) to the maximal one ( $H_o = 0.997$ ).



**Figure 23** Orientation analysis results (Part 2): karst networks with a clear vertical preferential direction. The karstic systems are classified from the ones with the lower orientation entropy ( $H_o = 0.918$ ) to the maximal one ( $H_o = 0.996$ ). As the vertical direction does not appear on a Rose diagram, orientation entropy ignores this preferential direction. It only expresses the existence of other preferential directions.