2011-11

On the concepts of complex networks to quantify the difficulty in finding the way out of labyrinths

Physica A, Amsterdam : Elsevier BV, v. 390, p. 4673-4683, Nov. 2011
http://www.producao.usp.br/handle/BDPI/50018

Physica A 390 (2011) 4673–4683

Contents lists available at SciVerse ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

On the concepts of complex networks to quantify the difficulty in finding the way out of labyrinths

D.R. Amancio *, O.N. Oliveira Jr., L. da F. Costa

Institute of Physics of São Carlos, University of São Paulo, P.O. Box 369, Postal Code 13560-970, São Carlos, São Paulo, Brazil

ARTICLE INFO

Article history:
Received 6 April 2011
Received in revised form 4 June 2011
Available online 21 July 2011

Keywords:
Labyrinths
Mazes
Complex networks
Non-supervised learning

ABSTRACT

Labyrinths have been a tradition and part of the imagination of the human kind for centuries, and were probably built either as a challenge to make it difficult for someone to find the way out, or for aesthetic purposes. They are conventionally classified according to the country they were built, to the style (Roman, classic and contemporary) or to the construction site. In this study, we show that labyrinths can be modeled as complex networks, whose metrics can be used to classify them in terms of their difficulty to find the way out. This is performed by calculating the absorption time, defined as the time it takes for a particle on an internal node to reach an output node through a random walk. The absorption time correlates well with the shortest paths and length of the networks, as expected, and has a very high correlation (Pearson coefficient of 0.97) with the betweenness, therefore allowing one to quantify the level of complexity of any labyrinth. It is shown that the conventional classification is inappropriate to distinguish between labyrinths, because some with very similar properties exist in different countries or were built in distinct time periods. A refined analysis in 77 famous labyrinths indicated that the majority were built for aesthetic purposes, with relatively small absorption times. Furthermore, with the expectation maximization algorithm, we could combine the complex network metrics to identify four clusters of labyrinths that differ in terms of density and shape.

1. Introduction

Since ancient times, human beings have been interested in the patterns found in nature, particularly those related to visual perception. With such interest, people began to conceive their own shapes as is the case of labyrinths (mazes) whose motivation is still a matter of debate [1]. There are basically two types of mazes: those composed of winding paths that lead from one or more input points to an output point, whose goal is to challenge and confuse the walker; and those that may be circular or square, built for aesthetic purposes or mystical motivation, as in the search for inner peace [1]. References and applications of mazes in societies and civilizations are numerous. They appeared frequently in religious rituals and mystical beliefs; in prehistoric times, it was believed that they helped to confuse evil spirits. In later times, mazes were associated with narrow, tortuous paths to bring humankind to God, usually symbolized by the center of the maze. Labyrinths used to be engraved in Greek coins, as a symbolic meaning. Also in Greece, mazes were often mentioned in folk tales, such as in the famous Theseus and Minotaur's myth [2], which tells the story of a maze whose goal was to keep the beast (the Minotaur) imprisoned. Labyrinths have also been mentioned several times by William Shakespeare in his writings and plays, including Gonzalo in 'The Tempest'[1] and Titania in 'A Midsummer Night's Dream'[2]. In more recent times, Umberto Eco [3] and

* Corresponding author.
E-mail addresses: diego.amancio@usp.br, diegoraphael@gmail.com (D.R. Amancio).

[1] 'The Tempest' is estimated to have been written in 1610–1611.

[2] 'A Midsummer Night's Dream' is estimated to have been written in 1594–1596.

0378-4371/$ – see front matter
doi:10.1016/j.physa.2011.06.079

Mark Danielewski [4] also mentioned mazes in their manuscripts. Many labyrinths remain both in buildings and on the floors of churches [1]. Their presence has extrapolated the boundaries of literature and religion, appearing even in public places[3] [1]. With computer and video games, they are used in various scenarios to make the game plots more complex.

Labyrinths appear in a wide variety of sizes, types, lengths, shapes, and complexity. They are categorized traditionally according to their geographical location (country), style which may be Roman, Classic and Contemporary, or the construction site (church, turf and others). As will be shown in this paper, this classification is insufficient to distinguish between different types of labyrinths. In order to provide a more discriminating strategy, here labyrinths are modeled using the concepts and methodologies of complex networks (CN) [5–9]. In particular, our model considers each point of bifurcation as a vertex and each possible path between these points as edges, so that each labyrinth is represented as a network. After the modeling process, we extract several complex networks metrics [8] to show that, indeed, traditional classification is inconsistent in the sense that different topologies coexist within clusters belonging to the same class. We also propose a method based on random walks [10] to categorize mazes according to the difficulty for someone to find the way out. As a result, we found that aesthetic mazes are usually simpler, according to our difficulty index.

This paper is organized as follows. First, we show in Section 3 how labyrinths were modeled as complex networks, and specify the metrics used to characterize them. Also included in Section 3 is a methodology to compare networks using projections into a two-dimensional plan, detection of outliers and discrimination by symmetry. We then investigate in Section 4 the relationship between traditional and CN-based classifications. Methods to quantify complexity in labyrinths are discussed, which are based on the so-called absorption time measure. Section 5 closes the paper.

2. Related works on using graphs to solve mazes

Graphs have been applied for both generating and solving mazes. For generation, one of the simplest methods is to use a depth-first search [11] in a graph. In this method, a predetermined rectangular grid is defined separating the cells. With this algorithm a subgraph from the grid is generated with the requirement that the graph formed is a spanning tree.[4] To generate labyrinths with such structure, the algorithm starts at a particular cell (or node) labeling it "exit". Then, for each non visited neighbor the corresponding wall is torn down, with the choice of the neighbor being random, and the process is repeated for the chosen neighbor. Finally, if there is no unvisited neighbor, the process simply backtracks.

Regarding the resolution of mazes, there are basically two types of algorithms: those using global information (which are more suitable when one has a schematic of the whole maze) and those using only the local information (more suitable for a walker inside an unknown maze). Among the algorithms based on global structure, we highlight the shortest path algorithm [11], which is especially useful if there are more than one solution. One example of efficient algorithm is the Dijkstra's algorithm [12], whose complexity is $O(n^2)$. Additional examples of algorithms are found in Ref. [11]. Another global approach is the algorithm referred to as dead-end filling, which basically converts a maze into a tree finding dead-ends and then connecting them to the closest bifurcations of the graph. Concerning the local strategy, one of the simplest strategies is known as wall follower, which ensures that if one keeps one hand in contact with the wall, then the way out will be found, provided that the representation of the maze as graph generates a connected network. Another local approach that tends to be more efficient is the Trémaux algorithm. Basically, it is a heuristic that requires the walker to draw lines on the floor to mark already visited paths. This method implements an informed depth first search – where visited paths are marked on the ground at most twice (when going or when returning) – so that the next neighbor to be chosen is the one with the fewest markings. When the way out is found, the most efficient solution is obtained by following the paths marked just once. It is also worth mentioning the algorithm normal strategy [13]. In this heuristic, non-traversed paths are chosen randomly while at nodes where there are no non-traversed paths, the latest path traversed only once is chosen. Interestingly, despite being very simple, this heuristic was the optimal solution in the maze problem formulated as a two person game: the setter and the solver [13].

In summary, algorithms have been proposed to analyze mazes from the point of view of graph theory, some of which focused on the problem of finding a solution to mazes. While in our work we are also interested in such solutions, our main focus is the investigation of how topology can be used to quantify complexity and classify mazes with similar structural features.

3. Methodology

3.1. Modeling labyrinths as networks

A database containing 77 images of mazes was collected from the Internet [14], which were initially classified – as is traditionally done – according to three features: country (Germany, France, Italy and England), style (Roman, Classical and Contemporary) and building site (church, rock, water and turf). To illustrate, Fig. 1 displays the scheme derived from the Hampton maze. In order to characterize mazes using concepts and methodologies of complex networks, each scheme

[3] Surprisingly, the number of public mazes seems to have increased since 1980, e.g. from 50 to 150 in Britain.

[4] Note that the graph generated must have a tree structure (i.e., cycles or disconnected components are not allowed). If the subgraph were disconnected, then there would be regions that would be wasted because they would not contribute to the search space. If the graph contained cycles, multiple paths between the chosen nodes would exist, thus facilitating the solution of the maze.

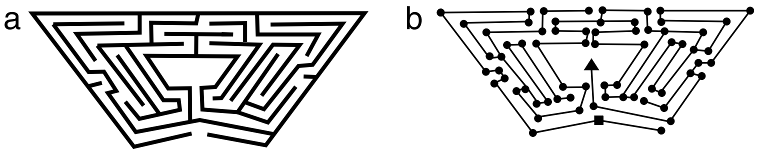

Fig. 1. (a) Schematic diagram of Hampton maze, constructed around the year 1690 by George London and Henry Wise at Hampton Court Palace in London, England. (b) Schematic diagram illustrating the complex network obtained. The nodes marked represent the input (triangle) and the output (square).

was transformed into a network with the help of homemade software specifically developed for this task. Basically, the semi-automatic process implemented transforms a maze in a network according to the following rules. Each bifurcation or inflection point is modeled as a node and the adjacency relationships between points that are not separated by barriers generate the set of edges. We chose to model the labyrinths taking the inflection points (even without bifurcation) because they represent decision points. For at an inflection point a person may decide that it would be more appropriate to return or continue in the walk. After the construction of the network, the input and output points are marked as they are used to quantify the difficulty to escape the maze (see absorption time). An example of network obtained for the Hampton maze is also depicted in Fig. 1. Note that the special points are either input (triangle) or output (square) nodes.

Several metrics defined in graph and complex networks theory were calculated to characterize the topological, dynamic and geometric features of the networks. These metrics are described below. Four topological metrics were used to characterize the vertex centrality. The simplest, known as degree, is defined by the numbers of neighbors. Specifically, if W_{ij} represents the weighted matrix of the network, then the vertex degree k_i is given by Eq. (1), while the network degree is obtained with Eq. (2). Within the context of the analysis of mazes, this measure may indicate that a node with large degree is a vertex in the maze with high probability of being visited.

$$k_i = \sum_{j=1}^{N} W_{ij} \tag{1}$$

$$k = \frac{1}{N} \sum_{i=1}^{N} k_i. \tag{2}$$

Another measure often used to study topology is the shortest path l, defined between two nodes i and j as the minimum cost d_{ij} necessary to achieve vertex i starting from vertex j. In network theory, the cost is the sum of the weights on the edges belonging to the path. For a topological study, d_{ij} is defined as the minimum number of steps that one must travel to reach vertex j starting from vertex i, since the distance between any two neighboring vertices is unitary. By specifying all d_{ij}, one can define the average minimum path according as:

$$l = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j>i} d_{ij}. \tag{3}$$

Similarly to the degrees, this measure has a logical relationship with the concept of complexity. Indeed, one should expect that complex labyrinths, in contrast to the simple ones, are characterized by large distances between any of their points of bifurcation. In calculating the complexity of the mazes, we could have just computed the distance between the entry–exit nodes. Instead, we chose to use all nodes in Eq. (3) because, as we shall show, this measure is employed to characterize the labyrinth topology as a whole, so that labyrinths with similar global topologies can be clustered together.

The third important feature is the assortativity Γ. A network is assortative ($\Gamma > 0$) if highly connected vertices tend to connect to other highly connected vertices. In contrast, it is disassortative ($\Gamma < 0$) if highly connected vertices tend to connect to vertices with low connectivity. This metric is important because it is able to characterize the nature of networks [15–17]. For instance, networks formed by communities (e.g. social networks) are assortative [17] while networks built from syntactic relations between words are disassortative [18]. To quantify the assortativity or the disassortativity in Eq. (4), one must first compute k_i for each node and the total number of edges M. Moreover, each weight must be normalized so that each strictly positive w_{ij} is replaced by $a_{ij} = 1$, since the assortativity does not consider the weight of the connections. In other words, to compute Γ, the adjacency matrix a_{ij} is employed instead of the weighted matrix w_{ij}. Although this measure has not been employed to quantify complexity in other contexts, it may still be useful for identifying regular regions in the network. For example, if the network obtained from the maze is assortative, then highly connected nodes are probably connected to other highly connected nodes. Hence, there may exist regions where the vertices are highly interconnected, giving rise to communities or "reservoirs" (in the context of disease transmission [15]). Since these regions make the route more complex (as a "prisoner" will spend longer time to escape), this measure may also be a quantifier of complexity.

$$\Gamma = \frac{\frac{1}{M}\sum_{j>i} k_i k_j a_{ij} - \left[\frac{1}{M}\sum_{j>i} \frac{1}{2}(k_i + k_j) a_{ij}\right]^2}{\frac{1}{M}\sum_{j>i} \frac{1}{2}(k_i^2 + k_j^2) a_{ij} - \left[\frac{1}{M}\sum_{j>i} \frac{1}{2}(k_i + k_j) a_{ij}\right]^2}. \tag{4}$$

Finally, the so-called betweenness centrality ϱ was employed. This metric may be understood as the flow of shortest paths passing through a given node. Therefore, the higher the betweenness of a node the greater is the probability of this node be reached through shortest paths. In order to quantify this metric, let $\sigma(i, v, j)$ be the number of shortest paths going from node i to node j passing through node v and let $\sigma(i, j)$ be the number of shortest paths passing through i and j. After calculating both values, the betweenness can be calculated either locally (for a particular node), as shown in Eq. (5), or globally (all nodes are considered), as shown in Eq. (6). As we shall show in the results, such measure is specially useful to characterize the difficulty to find the way out in labyrinths.

$$\varrho_v = \sum_i \sum_j \frac{\sigma(i, v, j)}{\sigma(i, j)} \tag{5}$$

$$\varrho = \frac{1}{N} \sum_{i=1}^{N} \varrho_v. \tag{6}$$

To characterize the dynamics of the network, two metrics were employed. The first, referred to as diversity [19–21], is defined as follows. Let Ω be the set comprising all nodes but the reference node i. So, the diversity for i can be obtained from Eqs. (7) and (8), where P_h is the probability that a random path is followed from i to j, passing exactly through h edges. Basically, this measure captures how different is the access from vertex i to other vertices when a random walk of length h (ranging from 1 to 3) is followed. In other words, this measure quantifies how many different positions can be effectively reached after h steps [19]. Thus, the complexity quantification may be done keeping in mind that high values of diversity indicates high levels of difficult in choosing the correct path. In addition to the interpretation related to the number of accessible neighbors, diversity metrics have proven suitable to detect borders in geographic networks [21].

$$\xi_h(\Omega, i) = -\frac{1}{\log(N-1)} \sum_{j=1}^{N} \pi(j, i) \tag{7}$$

$$\pi(j, i) = \begin{cases} P_h(j, i) \log(P_h(j, i)) & \text{if } P_h(j, i) \neq 0 \\ 0 & \text{if } P_h(j, i) = 0. \end{cases} \tag{8}$$

The network dynamics was also checked by calculating the absorption time τ [22], which quantifies how fast a randomly walking particle is absorbed by output vertices, assuming that the particle begins the random walk from input nodes. To calculate this metric, a Markov chain [23] was created with transition probability for a given output vertex i equals to $p_{ii} = 1$, since the particle does not return to the labyrinth after reaching the output. If node i is not an output, then the transition probability p_{ij} which considers the geometric feature of the network is given by:

$$p_{ij} = \frac{W_{ij}}{s_i}, \tag{9}$$

where s_i is the strength of node i:

$$s_i = \sum_{j=1}^{N} W_{ij}. \tag{10}$$

Let Θ be the submatrix representing the transitions between transient vertices (i.e., non absorbent vertices) and Φ be the submatrix depicting the transitions between transient vertices and absorbent vertices (i.e., output vertices). Then, arranging in a canonical form the matrix representing the transitions in the Markov chain, the matrix Π is obtained. Note that the third submatrix is the null matrix because there is no transition from absorbent nodes to transient nodes. Similarly, the fourth submatrix is the identity matrix because a particle that reaches an absorbent node stops walking.

$$\boldsymbol{\Pi} = \begin{pmatrix} \Theta & \Phi \\ 0 & I \end{pmatrix}.$$

To compute the absorption time, it is necessary to determine the time spent in transient nodes. For this, one first estimates the number of times that a particle starting from a given node i passes through a vertex j, before reaching an absorbent node. Let $\Psi = \psi_{ij}$ be the matrix that stores this information. We shall show that Ψ_{ij} can be calculated from Eq. (11):

$$(I - \Theta) \sum_{k=0}^{n} \Theta^k = I - \Theta^{n+1}. \tag{11}$$

Assuming that $\Psi = (I - \Theta)^{-1}$, we multiply both sides of Eq. (11) by Ψ, thus yielding Eq. (12):

$$\sum_{k=0}^{n} \Theta^k = \Psi(I - \Theta^{n+1}). \tag{12}$$

Since we do not know how many steps the particle walks to reach the exit, we consider $n \to \infty$ to take into account all the possibilities. Then, by letting n tend to infinity in both sides of Eq. (12), we obtain Ψ in Eq. (13). Note that the term Θ^{n+1} disappears because Markov chains composed of transient states with transition matrix Q satisfy the property $Q^N \to 0$, when $N \to \infty$. Now, to realize that Ψ defined in Eq. (13) quantifies the number of times that a given particle passes through vertex j before being absorbed, consider the random variable $X^{(k)}$, whose value is 1 if the particle is at node j after k steps, or 0 otherwise (see Eq. (14), where $\Theta^k = \theta_{ij}^{(k)}$). Since the expectancy $E(X^{(k)}) = \Theta_{ij}^{(k)}$ (Eqs. (15) and (16)), the expected number of visits to node j, starting from i respectively after a number of n or infinite steps, is given by Eqs. (17) and (18).

$$\Psi = \lim_{n\to\infty} \left(\sum_{k=0}^{n} \Theta^k \right) (I - \Theta^{n+1})^{-1} = \sum_{k=0}^{\infty} \Theta^k \tag{13}$$

$$p(X^{(k)} = t) = \begin{cases} \theta_{ij}^{(k)} & \text{if } t = 1 \\ 0 & \text{if } t = 0 \end{cases} \tag{14}$$

$$E(X^{(k)}) = 0 \cdot p(X^{(k)} = 0) + 1 \cdot p(X^{(k)} = 1) \tag{15}$$

$$E(X^{(k)}) = p(X^{(k)} = 1) = \Theta_{ij}^{(k)} \tag{16}$$

$$E\left(\sum_{k=0}^{n} X^{(k)} \right) = \sum_{k=0}^{n} \theta_{ij}^{(k)} \tag{17}$$

$$E\left(\sum_{k=0}^{\infty} X^{(k)} \right) = \sum_{k=0}^{\infty} \theta_{ij}^{(k)} = \psi_{ij}. \tag{18}$$

The absorption time can be obtained using Eq. (19), where the element τ_i of τ represents the absorption time for the node i and u represents a column vector with unitary elements. In fact, with such definition for u the element τ_i corresponds to the sum of the corresponding row of Ψ, as shown in Eq. (20).

$$\tau = \Psi u \tag{19}$$

$$\tau_i = \sum_{j=1}^{N} \Psi_{ij}. \tag{20}$$

We chose this measure to characterize the difficulty in finding the way out of labyrinths because a random walk is one of the most naive possible heuristics to walk through the labyrinth. As could be noticed in the derivation of the measure, no information on the visited nodes is taken into account. For this reason, random walks estimate the maximum expected time (worst scenario) for finding the way out, when compared to clever strategies. In fact, more efficient algorithms which mark previous visited nodes [13] could lead to shorter absorption times, but for the purpose of our paper – which is to compare different mazes – taking the worst case is still valid.

As mentioned before, an algorithm using information about the already visited vertices nodes tends to be more efficient than the strategy based on random walks. To verify the relationship between these non informed and informed searches, we also tested an algorithm based on an informed depth-first search [11,24]. In this heuristic the walker walks in the network identifying the points already visited. When a dead end (i.e., a node whose all neighbors have been already visited) is reached, the walker returns to the previous node marked in the list of visited nodes. This walk continues until one of the outputs is found. Finally, the complexity of the maze is computed as the total traveled distance. Since it is necessary to pick randomly the next neighbor to be visited if more than one neighbors have not as yet been visited, the complexity computation may vary among different executions of the heuristic. For this reason, we run the heuristic 300 times for each maze and considered as measure of complexity the average complexity over these 300 observations.

3.2. Additional metrics and visualization

The characterization was complemented by computing some geometric metrics. They are called geometric because they take into account not only the relations between vertices, but the distance of these relations.The reason why we also employed geometric features was to take into account the easiness with which one can move between two points. If these geometric features were not considered, close and distant points would be taken as having the same difficulty in their routes. Basically, in this case the shortest paths were computed, considering the minimum distance between adjacent vertices as the cost function.[5] Also, the sum of all edges weights (i.e. the network length) was exploited in the experiments.

The set of metrics obtained with the labyrinths modeled as complex networks may be combined with statistical tools for the purpose of classification. Here we used principal component analysis (PCA) [25,26] to project the network data on a

[5] In order to homogenize the distances, they were normalized according to the size of each image.

2D plot. This transformation was used to eliminate possible correlations through a new coordinate system whose principal coordinate accumulates the largest variances of the data. To project each vector $\overrightarrow{v}$ describing a labyrinth, the covariance matrix Σ is calculated:

$$\Sigma(i,j) = \frac{1}{\eta - 1} \sum_{k=1}^{\eta} (v_k(i) - \mu_i)(v_k(j) - \mu_j), \tag{21}$$

where $v_k(i)$ is the kth-element of the vector describing the measurement i values and μ_i is the average of $v_k(i)$ over the η labyrinths, i.e.

$$\mu_i = \frac{1}{\eta} \sum_{k=1}^{\eta} v_k(i). \tag{22}$$

From Σ, one calculates the eigenvectors related to the eigenvalues with the largest absolute values. These eigenvectors are used to define the two-dimensional plane, where each vector $\overrightarrow{v}$ will be projected. After such processing transformations, supervised and unsupervised machine learning algorithms were applied to find similarities between the labyrinths studied.

3.3. *Measuring symmetry*

To quantify the symmetry for each maze, two procedures were followed, as illustrated in figure S1 in the Supporting Information. First, the color image characterized by three matrices (red (R), green (G) and blue (B)) was converted to a gray level image using the following expression [27]:

$$Y = 0.3 \cdot R + 0.59 \cdot G + 0.11 \cdot B \tag{23}$$

where Y is the matrix representing the image. Note that even though the conversion of an image to grayscale is not unique, the values of 30% of the red value, 59% of the green value, and 11% of the blue value are typically used [27] as they are related to visual sensitivity of the human eye to the conventional primary colors. After conversion, the symmetry S was quantified according to Eqs. (24) and (25), which compare pixels positioned at the same distance from the vertical symmetry axis. Despite the possible existence of several axes of symmetry, the value of S quantified here refers only to the vertical axis.

$$S = \frac{1}{N\lfloor M/2 \rfloor} \sum_{i=1}^{N} \sum_{j=1}^{\lfloor M/2 \rfloor} \delta(y(i,j), y(M - j + 1)) \tag{24}$$

$$\delta(i,j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise .} \end{cases} \tag{25}$$

3.4. *Detecting outliers*

In statistical studies, outliers are defined as instances that are widely dissimilar from other observed instances. In particular, during this study we looked for mazes with dissimilar formats to verify the influence of non conventional shapes to the topology of the corresponding networks. The method adopted to detect such outliers is described as follows. In the first step, each observation characterized by the set of measures extracted from the networks was projected using the procedure described in Section 3.2, as illustrated in the left panel of Figure S2 in the Supporting Information, where we considered the problem of finding the outliers of a given database comprising 9 observations. Then, for each point p represented in the plane, we associated a binary relation $f_p(r)$ that counts the number of points that lie at a distance less than r, where r ranges from 0 to the greatest potential distance r_m, which depend on p. By way of illustration, Figure S2 of the Supporting Information displays in the central and in the right panel the behavior of $f_p(r)$ for the points marked as asterisk and square, respectively.

After defining f, let us define $o(p)$ with Eq. (26), where N is the number of points projected into the plan. Since $f_p(r)$ reaches the maximum value when $r \simeq r_m$ if p is an outlier, the values of $o(p)$ tend to be very small (see example in Figure S2 of the Supporting Information). For this reason, we infer that p is an outlier when $o(p)$ is greater than an arbitrary threshold value α, which ranges from 0 to 1.

$$o(p) = \frac{1}{(N-1)r_m} \int_0^{r_m} f(r)\mathrm{d}r. \tag{26}$$

4. Results and discussion

4.1. *Classification of labyrinths using non-supervised learning methods*

We examined the hypothesis that metrics extracted from complex networks can cluster similar labyrinths. Clusters were generated using topological, geometric and dynamics metrics, with classifiers induced by the expectation maximization algorithm [28]. This algorithm was chosen because it does not require prior specification of the number of clusters, since

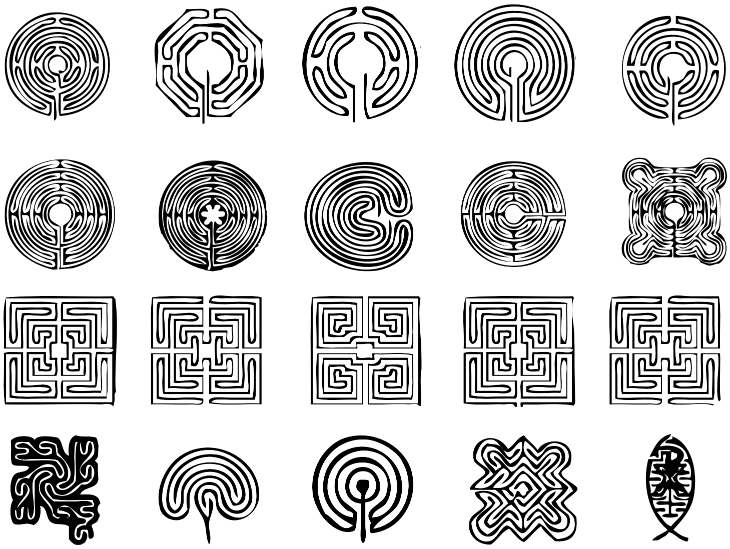

Fig. 2. Representative labyrinths of the 4 clusters identified using geometric metrics with the expectation maximization algorithm. Each row brings some of the labyrinths of a given cluster.

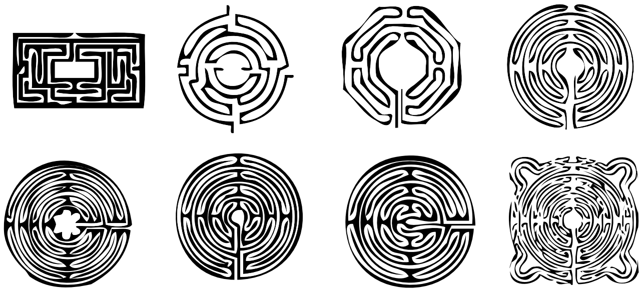

Fig. 3. Two clusters of labyrinths identified when all metrics (geometric, dynamics and topological) were used in conjunction with the expectation maximization algorithm.

this information is initially unknown. The best result was obtained when only geometric metrics are employed. In this case, four groups were obtained, as illustrated in Fig. 2, where each row represents one of the four clusters. As shown, the groups are fairly distinguishable in the following classes: short mazes (first row), long mazes (second row), regular mazes (third row) and irregular mazes (fourth row). In fact, the clusters seem to vary in complexity, as will be quantified later on when the absorption times are introduced.

When all metrics are used together with the same inductor algorithm, two clusters were obtained as depicted in Fig. 3. The visual distinction is not as good as that in Fig. 2, but it is still possible to observe dense and sparse labyrinths. Thus, the cluster analysis in both figures suggests a new taxonomy for the labyrinths, based on regularity and density. Moreover, since some clusters (e.g., the first and the second in Fig. 2) display different degrees of complexity, it is possible to conclude that our models are suitable to quantify such a degree.

The analysis with complex networks indicates that the conventional ways to classify labyrinths according to the country, style or building site where they appear, is not appropriate to distinguish among the various labyrinths in terms of their structural properties. That is to say, labyrinths from distinct countries may be very similar to each other, while differing from other labyrinths of the same country. The same applies to the style or environment. This is not surprising if one considers the history of labyrinths, as in some cases they were designed by the same architect and reproduced in different countries, or were inspired in existing labyrinths from other time periods. Analytically, these statements are shown in the dendrograms obtained by hierarchical clustering using the Wards distance [29] in Fig. 4 (see also Figures S3 and Figures S4 in the Supporting Information). Although in some cases the classification coincides reasonably well with the conventional classification (e.g., Roman style in Fig. 4), the overall agreement is poor.

The results above clearly indicate that geographical and cultural factors are not appropriate to distinguish between different types of labyrinths, since the traditional classification does not take into account topological features, such as complexity. One could perhaps recall that many labyrinths were conceived to be difficult for someone to find the way out

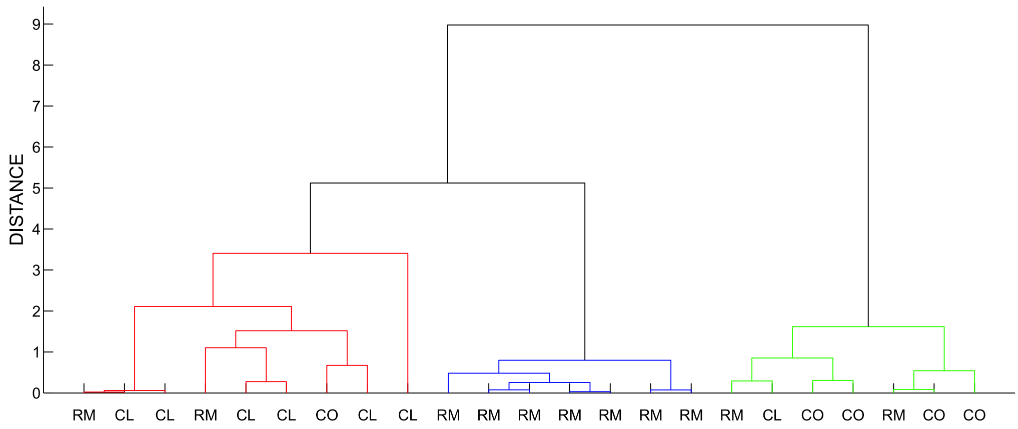

Fig. 4. Hierarchical clustering using the wards distance for the attributes obtained with the 2D projection using PCA. Three styles were considered: Roman (RM), Classic (CL) and contemporary (CO). Only the second cluster had the predominance of a style (RM).

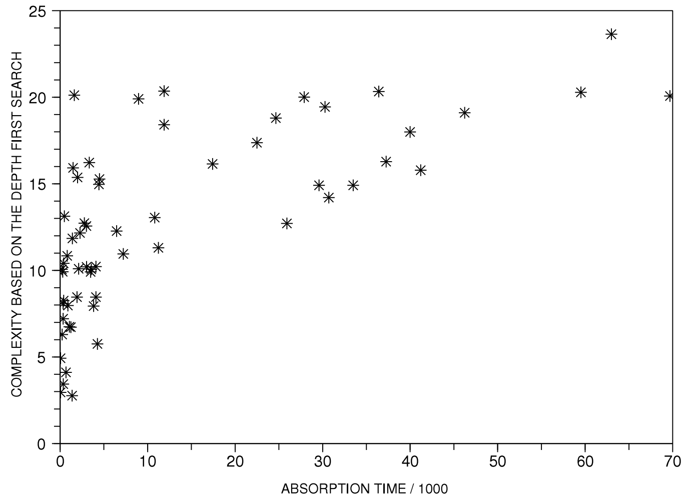

Fig. 5. Relationship between absorption time (based on a not informed search) and complexity calculated using an informed depth-first search. The high value of the Pearson coefficient ($r = 0.69$) indicates that the two approaches are similar for comparing mazes with regard to their complexity, even though the times are higher for the random walk as one should expect.

(e.g., in the myth of Ariadne's Thread, the character Daedalus built a house confusing the usual passages with a maze of various winding paths). Since complexity seems to be a more suitable way to classify labyrinths, in the next section we introduce an approach to quantify such difficulty.

4.2. Quantifying labyrinth complexity

We suggest here that the difficulty in finding the way out from a labyrinth is related to the time it would take for a particle (starting from the input node) to reach (or be absorbed) an output node. To estimate this time, we used the two strategies described in Section 3.1: the absorption time in a random walk and the time based on an informed depth first search. A strong correlation between the two methods was found ($r = 0.69$), as illustrated in Fig. 5. This means that although the informed strategy is more effective than the one based on random walks, the relationship between the complexity provided by both strategies is roughly linear. For this reason, in the following analysis, we consider only the absorption time as a quantifier of complexity.

The labyrinths exhibiting the longest and shortest absorption times are shown in Fig. 6. A visual inspection confirms one's intuition that longer absorption times should be observed for noticeably more complex labyrinths. Also interesting is to observe that some labyrinths appear not to be meant to be difficult, as in Fig. 6 (a) there are cases of a direct path between the starting and ending points. In fact, in addition to the difficulty in finding the way out, another criterion used in conceiving labyrinths was the beauty and symmetry of the design. In order to further explore this point, each of the authors in this paper chose independently (by visual inspection) mazes considered aesthetic or with symbolic meaning. For the mazes for which all authors agreed to their aesthetic nature, the absorption time ranged from 120 to 3000, to be compared to the longer

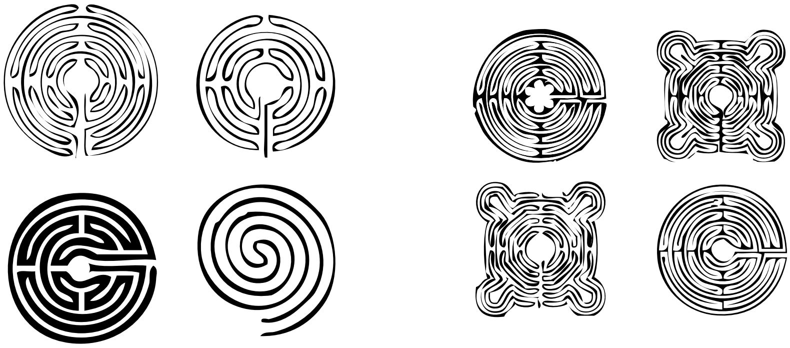

Fig. 6. The four labyrinths with the smallest (on the left) and with the largest (on the right) absorption times. Note that while it is easy to find the way out of the labyrinths with small absorption times because the distances between input and output nodes are small.

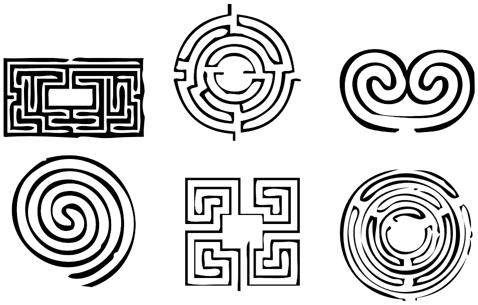

Fig. 7. Examples of aesthetic labyrinths, which normally have short absorption times.

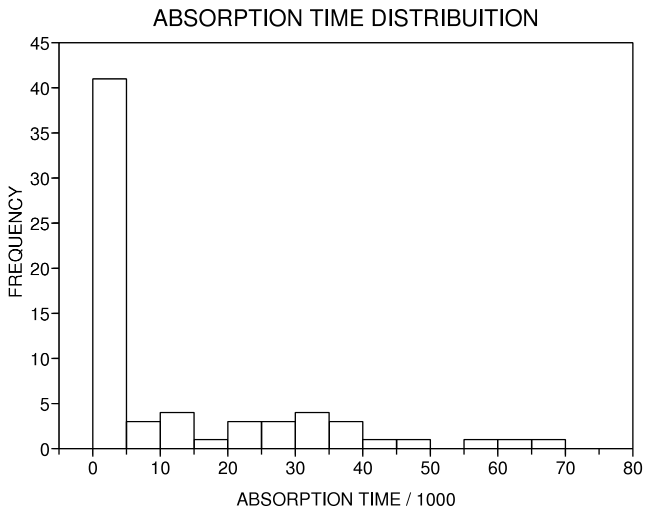

Fig. 8. The distribution of absorption times reveals that only a small fraction of labyrinths have high absorption times.

absorption time of 69,600 when all the 77 labyrinths were analyzed. Some of the labyrinths considered aesthetic are shown in Fig. 7. Still with regard to the motivation for building the labyrinths, about 70% of the labyrinths analyzed had short absorption times (see distribution in Fig. 8). This means that the idea of building simple, aesthetic labyrinths appeared to prevail over the idea of making difficult designs.

Upon examining the correlation between absorption time (i.e. difficulty to find the way out) and other measures using the Pearson correlation coefficient [30], we note in Fig. 9 that complexity correlates well with the shortest paths and length, which should be expected. What was less intuitive was the high correlation with betweenness, with only a few labyrinths deviating from this behavior (see examples in Figure S5 of the Supporting Information). Therefore, the more the network is overloaded (i.e., the greater the flow passing through the vertices) the higher its complexity.

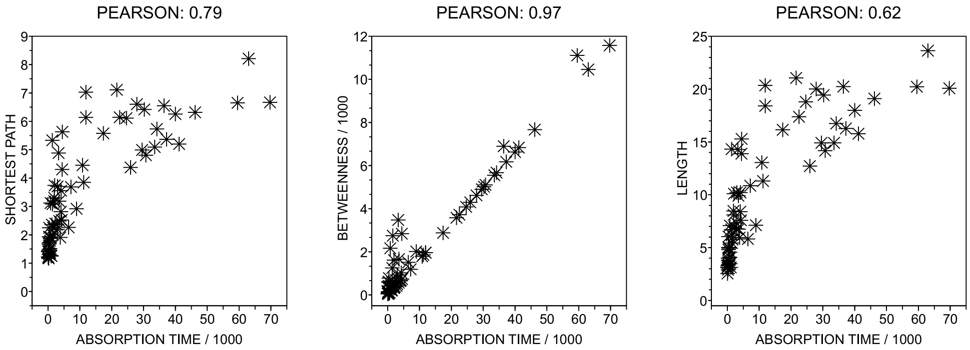

Fig. 9. Scatter plots showing the correlation between absorption time and other measures. Note in particular that the Pearson coefficient correlation is close to 1 for the betweenness. Therefore, the absorption time is strongly related to the network overload.

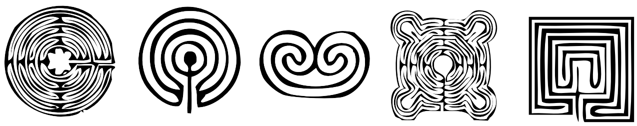

Fig. 10. Outliers found using the method based on principal components analysis. Indeed, there exists a strong correlation between shape and complexity, since labyrinths classified as outliers are either very trivial or very complex.

We have also verified that the conventional classification for the labyrinths (country, style or construction site) is not appropriate to determine the level of difficulty to find the way out. Indeed, when the absorption time was used in a hierarchical classifier with the Wards distance, the distinction was poor (see Figure S5 of the Supporting Information) for the styles. This also applied to the classification by country and construction site (results not shown).

4.3. Outliers and absorption time

Although uniform standards are important to characterize and classify real systems, one must pay attention to deviations from the mean, as they can provide useful information about the particularities of each system. For this paper it is reasonable to say that a labyrinth is an outlier if its characteristics (described by the networks measures) are topologically irregular when compared to the others. Using the methodology described in Section 3.4 and all the metrics (dynamics, topological, geometrical), we identified outliers in the data set. With $\alpha = 0.3$, the outliers displayed in Fig. 10 were obtained, which are quite distinct from the remainder ones, in terms of both the density and complexity. Indeed, 3 out of the 5 outliers had the highest absorption times while the remaining had low absorption times (among the 20 mazes with lowest absorption times).

The combination of topological and geometrical measures to characterize the networks and detect outliers showed that the structural characteristics of the network can also be used to assess maze complexity. In fact, outliers were shown to represent either complex or simple labyrinths. It is therefore possible to employ a multivariate strategy to classify mazes according to complexity. In particular, one could say with higher accuracy that a given maze is complex or not when it is classified as outlier, since all measurements (not only the absorption time) are employed to reveal an unusual structure.

5. Conclusion

The characterization of labyrinths represented as networks and using various metrics has shown that the traditional classification, in terms of country, style and building site, is not appropriate to distinguish their topological and geometrical features. Indeed, labyrinths with very similar properties could have been created in different countries or in different sites. Through exploring several metrics and using machine learning methods, we managed to obtain classes of labyrinths that are distinguishable in their complexity. To quantify the degree of complexity, we proposed the measurement referred to as absorption time, taken as indicative of the difficulty in finding the way out from a labyrinth. In terms of other complex network properties, the absorption time was correlated with the length and shortest paths, as it should be expected, and with the betweenness centrality. Also interesting was the finding that the majority of the labyrinths were not meant to be complex. Approximately 70 % of them displayed small absorption times, which means that the construction with aesthetic purposes prevailed over the intention of making a complicated path. Finally, we detected the labyrinths considered as outliers to show that topological and dynamical measurements can be combined to create a more powerful method to quantify the complexity to find the way out, since outliers represent either complex or aesthetic labyrinths.

We believe that the results reported here will not only allow for a new taxonomy in the study of labyrinths, but also provide a useful tool to quantify complexity in other applications, such as computer games. Furthermore, the generic methods employed here may be applied to other network problems, since any network can be thought of as a set of sinuous paths, i.e. a labyrinth.

Acknowledgments

This work was supported by the FAPESP (proc. 2010/00927-9 and 2005/00587-5) and the CNPq (proc. 301303/06-1). The authors are thankful to Mr. Barry Timmons for valuable discussions.

References

[1] J. Saward, Magical Paths: labyrinths and mazes in the 21st century, Mitch, 2008.
[2] P. Conty, The genesis and geometry of the labyrinth; architecture, hidden language, myths, and rituals, 2002.
[3] U. Eco, The Name of the Rose, first ed., Harvest Books, 1994.
[4] M.Z. Danielewski, House of Leaves, second ed., Pantheon, 2000.
[5] M.E.J. Newman, The structure and function of complex networks, SIAM Review 45 (2003) 167–256.
[6] A.L. Barabasi, E. Bonabeau, Scale-free Networks, Scientific American, 2003.
[7] M. Newman, A.L. Barabasi, D.J. Watts, The Structure and Dynamics of Networks, in: Princeton Studies in Complexity, 2006.
[8] L F Costa, F A Rodrigues, G Travieso, P R Villas Boas, Characterization of complex networks: a survey of measurements, Advances in Physics 56 (2007) 167–242.
[9] L.F. Costa, et al., Advances in Physics 60 (2011) 329–412, as the paper originally in arxiv has now been published.
[10] B. Luque, O. Miramontes, Small-worlds, mazes and random walks, Europhysics Letters 63 (1) (2003).
[11] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, Introduction to Algorithms, MIT Press and McGraw-Hill, 2001.
[12] E.W. Dijkstra, A note on two problems in connexion with graphs, Numerische Mathematik 1 (1959) 269–271.
[13] E.J. Anderson, Mazes: search games on unknown networks, Networks (1981) 193.
[14] The images of labyrinths displayed in this paper were adapted from the images available in the following websites (date of the retrieval: June/2011), http://en.wikipedia.org/wiki/File:Labyrinth_1_%28from_Nordisk_familjebok%29.svg, http://en.wikipedia.org/wiki/File:Hampton_court_maze.svg, http://www.genese.com/labirinto, http://www.u-picardie.fr/patrick/Cathedrale/Interieur/nef/Labyrinthe.html, http://gwydir.demon.co.uk/jo/maze, http://www.labyrinthos.net, http://www.jeu-labyrinthe.com All labyrinths can also be directly downloaded from http://dl.dropbox.com/u/2740286/labyrinths.zip.
[15] M.E.J. Newman, Assortative mixing in networks, Phys. Rev. Lett. 89 (2002) 208701.
[16] M.E.J Newman, Finding community structure in networks using the eigenvectors of matrices, Physical Review E 74 (2006) 036104.
[17] M.E.J. Newman, Mixing patterns in networks, Physical Review E 67 (2003) 026126.
[18] R. Ferrer i Cancho, R.V. Solé, R. Köhler, Patterns in syntactic dependency networks, Physical Review E 69 (2004) 051915.
[19] L. Antiqueira, F.A. Rodrigues, L.F. Costa, Modeling connectivity in terms of network activity, Journal of Statistical Mechanics (2009) L09005.
[20] F.A. Rodrigues, L.F. Costa, Comparison of the interactomic networks of different species in terms of accessibility, Molecular Biosystems 6 (2009) 224.
[21] B.A.N. Travençolo, M.P. Viana, L.F. Costa, Border detection in complex networks, New Journal of Physics 11 (2009) 063019.
[22] J.L. Katz, R.L. Burford, Estimating mean time to absorption for a markov chain with limited information, Applied Stochastic Models and Data Analysis 4 (4) (2006) 217–230.
[23] J.R. Norris, Markov Chains, Cambridge Series in Statistical and Probabilistic Mathematics (1998).
[24] M.T. Goodrich, Algorithm Design: Foundations, Analysis, and Internet Examples, Wiley, 2001.
[25] L.F. Costa, R.M. Cesar Jr., Shape Analysis and Classification: Theory and Practice, CRC Press, 2009.
[26] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, JohnWiley and Sons, 2001.
[27] W.K. Pratt, Digital Image Processing, John Wiley and Sons, New York, 1991.
[28] D.J.C. MacKay, Information Theory, Inference and Learning Algorithms, first ed., Cambridge University Press, 2002.
[29] J.H. Ward, Hierarchical grouping to optimize an objective function, Journal of the American Statistical Association 58 (301) (1963) 236–244.
[30] J. Neter, M.H. Kutner, C.J. Nachtsheim, W. Wasserman, Applied Linear Statistical Models (1996).