Available online at www.sciencedirect.com
ScienceDirect
Physica D 224 (2006) 149–155
PHYSICA D
www.elsevier.com/locate/physd

Analysis of attachment models for directory and file trees

Konstantin Klemm[a], Víctor M. Eguíluz[b,*], Maxi San Miguel[b]

[a] *Bioinformatics, Department of Computer Science, University Leipzig, Härtelstr. 16-18, 04107 Leipzig, Germany*
[b] *Instituto Mediterráneo de Estudios Avanzados IMEDEA (CSIC-UIB), E07071 Palma de Mallorca, Spain*[1]

Available online 3 November 2006

Abstract

Many networks emerge as the outcome of a collective interaction, such as the World Wide Web (WWW); others are the consequence of the biological evolution, such as the brain. In contrast to these examples, we investigate the topology of trees generated by single individuals. Computer users generate directory structures to store and manage information in files. Analyzing the directory and file trees generated by different users we have access to different realizations available for statistical analysis. We characterize the architecture of directories and files created by different computer users by means of the degree distributions and number of leaves, degree–degree correlations, average distance to root, and community size distributions. We compare the different topologies in the search for similar managing patterns, and compare the trees obtained with two simple models of growing networks and with a model that interpolates between them and incorporates the heterogeneity of the computer users.

Keywords: Directory trees; File trees; Complex networks

1. Introduction

It is superfluous today to emphasize the importance of the processes of storing and retrieving information in the new knowledge-based society. An interdisciplinary area of research dealing with information and knowledge management has emerged and is being termed 'mapping knowledge domains' [1]. Questions considered in this framework address the way in which information is stored, organized and retrieved. A general goal is to improve the work of search engines that currently only index a small fraction of the WWW [2].

In this context we analyze in this paper different possible models to describe the tree structure of the files stored in a computer cluster by the users of the computer facilities of our own Department. In comparison with other studies in the broad literature on data analysis of complex networks we identify two specificities of our data. First, the networks analyzed are not the result of a collective action of agents, but something created by a single individual. Secondly, we can analyze statistically different realizations of different sizes of the network, since each user of the computer cluster has created its own tree with a different number of files.

Possible extreme models for the structure of the file tree that we analyze are one describing a random process of file-storing or the alternative one that describes a fully rational and optimized process of file storing. Our data is best described by models that incorporate randomness and arbitrary choices within rational design: in the framework of the discussion of tinkering versus engineering in natural or technological realities [3,4] we might say that what we produce while storing our files is an artificial reality by a tinkering process similar to the way that complex natural systems are believed to operate. The old-style good engineer works according to a preconceived plan aiming to deliver an 'ordered' perfect object. However, the new complex technological realities are the result of a flexible, self-organizing and adaptable process, like the WWW, or the result of an engineering design that has to allow for functionality, evolution, changing environment, reuse and refactoring, as, for example, software systems [5,

* Corresponding author.
E-mail addresses: klemm@bioinf.uni-leipzig.de (K. Klemm), victor@imedea.uib.es (V.M. Eguíluz), maxi@imedea.uib.es (M. San Miguel).
URLs: http://www.bioinf.uni-leipzig.de/~klemm/ (K. Klemm), http://www.imedea.uib.es/~victor (V.M. Eguíluz), http://www.imedea.uib.es/~maxi (M. San Miguel).
[1] http://www.imedea.uib.es/physdept.

0167-2789/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physd.2006.09.033

6]. Therefore, these artificial realities are to a great extent the result of tinkering, with building mechanisms that appear to be similar to those naturally used in the management of our own knowledge domain.

An interesting question is to decide if the structure of the file trees belongs to a class of efficient networks, that is, if it can be obtained from some optimization principle. River networks [7,8] and the vascular system [7,9,10] are examples of efficient networks in which transportation costs are minimized. This is captured in the *allometric scaling* relating topological properties with network size. For instance, the binary tree of communities [11] obtained from the e-mail network of a real organization [12] seems to belong to the same class as the river networks, for our file trees we find exponents in the same class as food webs [13] and the community of scientific collaboration [12]. It is also interesting to note that the optimization principle defining the efficient networks discussed in Ref. [7] is not the only possible optimization principle. In fact, it has been shown that a different optimization process can explain the selection of preferential attachment strategies [14]: by tuning a parameter which weights two contributions in the optimizing energy, preferential attachment appears at the boundary between random and forced attachment.

In this paper we provide an extensive characterization of individual user computer directory and file trees, calculating a number of quantitative measures. In a first short report on this analysis [16] we gave some evidence that a model with a single parameter that balances the ratio between preferential and random attachment reproduces topological features of the directory trees. Here we include file and directory trees and we substantiate the merit of that model by comparing in detail our data with the predictions of a preferential attachment growing tree model and a random growing tree model. Although the pure preferential attachment growing model reproduces main features of the data, this comparison sets its limits of validity and elucidates the important role of the parameter introduced in the model of Ref. [16]. This parameter incorporates the heterogeneity of the individuals in the parameter independent preferential attachment mechanism. Our detailed comparison with the random and preferential attachment models includes the analysis of degree distributions (Section 2), degree–degree correlations (Section 3), average distance between files [17–20] (Section 4), and distribution of community sizes in the tree [12] and allometric scaling exponents [7,13] (Section 5). General conclusions and open questions are summarized in Section 6.

2. Degree distributions

We obtained data from 63 computer users of an academic research facility. The users range from permanent staff to temporal visitors, and include researchers from both sexes, and several ages and nationalities. The nodes in a directory tree are the directories or folders in a user's computer account, where two directories are connected if one of the directories is a subdirectory of the other. Two examples of trees in the data set are shown in Fig. 1. In addition to the directory trees we analyze *file trees*. In the latter, also the files in a user's account

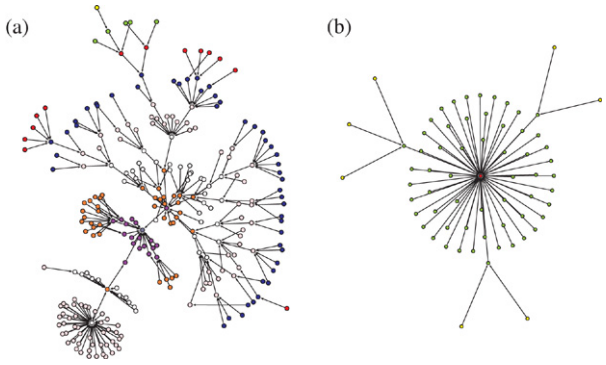

Fig. 1. Two directory trees of sizes (a) $N = 260$ and (b) $N = 66$.

are included as nodes. Each such node has a single link to the node that represents the directory in which the file is stored.

The degree k_i of a node i is the total number of links connecting i with other nodes. In a tree, the mean degree averaged over all nodes only depends on the number of nodes N. It is given by

$$\langle k \rangle = \frac{2(N-1)}{N} = 2 - \frac{1}{N}, \tag{1}$$

which approaches $\langle k \rangle \approx 2$ for large N. The simplest characterization of a network is its *degree distribution* $P(k)$, i.e. the fraction of nodes with degree k. In the analysis of empirical data, it is more convenient to calculate the cumulative degree distribution (especially for skewed distributions) given by

$$Q(k) = \sum_{j=k}^{\infty} P(j). \tag{2}$$

In other words, the cumulative distribution is the fraction of nodes with degree k or higher.

In Fig. 2 we show the cumulative degree distributions for the nine largest of the 63 directory and file trees. In all cases we observe distributions with a power law decay $Q(k) \sim k^{-\gamma+1}$ with an exponent in the range $2.2 < \gamma < 2.8$. The power law decay extends to the value of $k_{\max}$ such that $Q(k_{\max}) = N^{-1}$. This indicates that the cut-off of the distributions is only due to the finite size of the tree. There is no indication of an upper bound on the degree causing an explicit cut-off. Even though, the cumulative degree distributions of the file trees are not as good as for the directory trees, they are also skewed and show tails of power law decay (Fig. 2).

In order to understand the emergence of the degree distribution let us now consider simple models for the construction of a directory tree. We assume that users build their trees by iteratively adding nodes, *i.e.* creating new directories. The effect of eventual removals of nodes from the trees is considered to be negligible. Then we are left to defining a rule for the *attachment* of a new node. In Ref. [16], we introduced a model of a growing tree with a parameter that balances the ratio between preferential and random attachment. This model has been shown to capture the topology of the directory trees selecting a value of the parameter to each user. Here we focus on two limiting cases: pure preferential attachment and pure random attachment. In a first approach we assume that the

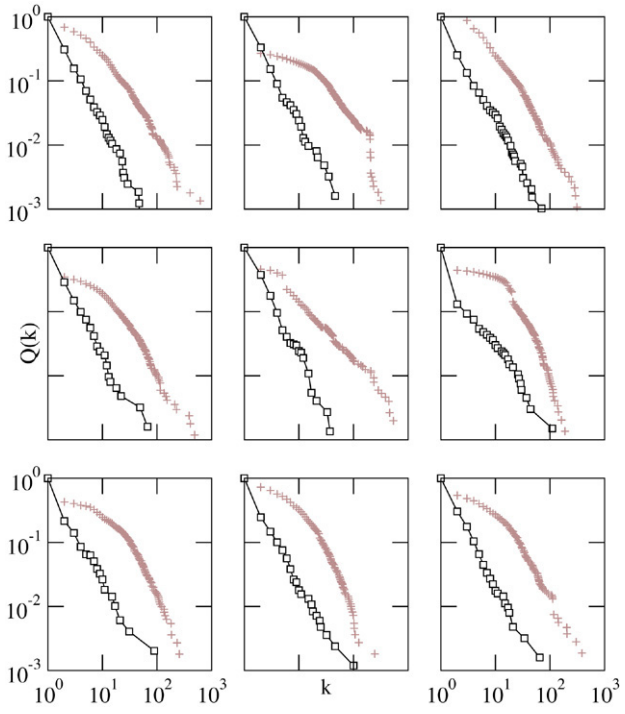

Fig. 2. Cumulative degree distributions (□) of the nine largest directory trees in the data set. For each of the systems also the cumulative degree distribution of the corresponding file tree is shown (+). The latter values have been multiplied by a factor of 10 for plotting. Sizes of these nine directory trees are in the range $500 < N < 1600$, for the file trees we have $10{,}000 < M < 22{,}000$.

user places a new directory completely at random: each of the existing directories has an equal probability of being chosen as a parent directory. In this *homogeneous attachment* model the cumulative degree distribution converges towards the geometric

$$Q^{\mathrm{hom}}(k) = \left(\frac{1}{2}\right)^{k-1}, \tag{3}$$

for $k \geq 1$, and $Q^{\mathrm{hom}}(0) = 1$. For a systematic comparison between the model and all 63 trees let us consider two derived quantities: the fraction of leaves and the largest degree. One is the fraction of leaves $P(1)$, the number of nodes with only a link. For the homogeneous attachment model we find

$$P^{\mathrm{hom}}(1) = Q^{\mathrm{hom}}(1) - Q^{\mathrm{hom}}(2) = \frac{1}{2}. \tag{4}$$

This is clearly below the values found for the empirical trees, cf. Fig. 3(a). There is also a clear discrepancy between the model and the data with respect to the largest degree k_{max} of a given tree. While empirically k_{max} grows algebraically (with an approximate exponent of $1/2$), the model yields

$$k^{\mathrm{hom}}_{\mathrm{max}} = \log_2 N \tag{5}$$

as a solution of the equation $Q^{\mathrm{hom}}(k^{\mathrm{hom}}_{\mathrm{max}}) = N^{-1}$. In summary, the homogeneous attachment cannot serve as a model for the construction of the directory trees because it does not reproduce the feature of the broad degree distribution.

In the context of complex networks, a model has been proposed that reproduces the power law distributed degree. In

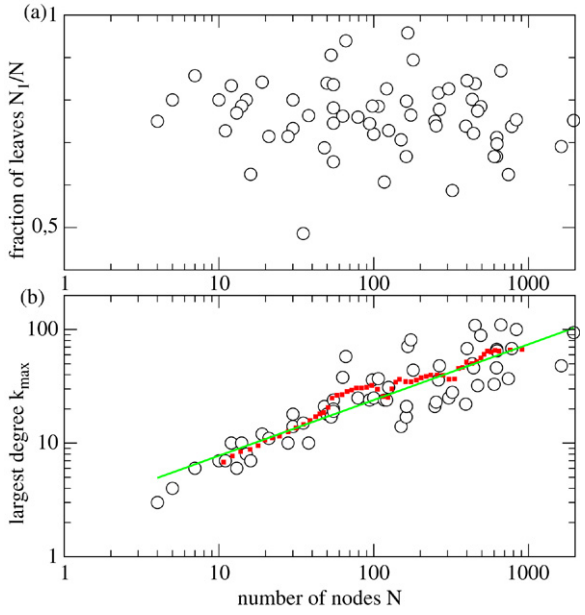

Fig. 3. (a) Relation between total number of nodes and the fraction of leaves for the 63 directory trees. With one exception, for trees of all sizes the fraction of leaves is clearly above $1/2$. (b) The largest degree k_{max} as a function of system size (circles). The squares are a running average with a window size of 10 data points. The solid line follows $k_{\mathrm{max}} \propto N^{0.49}$ as the result of a fit to the data.

this model, nodes are added iteratively to the existing structure just as in the above scenario. The target nodes to which they attach, however, are not chosen with equal probability. Linear *preferential attachment* is used: the probability that an existing node i receives the link from the newly added node is proportional to the degree k_i. For this model the degree distribution asymptotically decays as a power law

$$Q^{\mathrm{pref}}(k) \sim k^{-\gamma+1} \tag{6}$$

with $\gamma = 3$ [21].

As a technical detail, we note that the root node is assigned an extra 'dangling' link. Otherwise the attachment rule would not be defined in the case $N = 1$. Then the sum of the degrees of all N nodes in the system is $2N - 1$.

In this model the expected number of leaves N_1 grows as

$$\Delta N_1 = 1 - \frac{N_1}{2N - 1} \tag{7}$$

because each increment of system size $N \to N + 1$ adds a new leaf (first term) and each existing leaf receives a second link with probability $(2N - 1)^{-1}$ (second term). For large $2N \gg 1$, Eq. (7) is solved by $N_1 = (2/3)N$. The fraction of leaves is given by

$$P^{\mathrm{pref}}(1) = \frac{N_1}{N} = \frac{2}{3} \tag{8}$$

in the preferential attachment model. Compared with the homogeneous attachment model, the preferential attachment generates a larger number of leaves in trees of the same size. The latter model agrees better with the empirical data in Fig. 3(a).

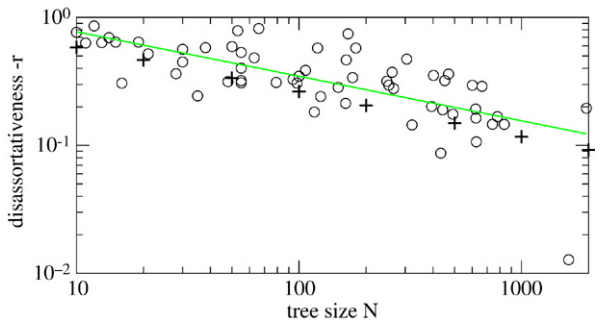

Fig. 4. Degree–degree correlation coefficient r. The solid line is the result of a power-law fit to the directory trees (○). The degree–degree correlation coefficient is also plotted for trees generated with the preferential attachment model (+).

The expected largest degree is subject to the growth equation

$$k_{\max}^{\mathrm{pref}}(N+1) = \left[1 + \frac{1}{2N-1}\right] k_{\max}^{\mathrm{pref}}(N) \tag{9}$$

with the asymptotic solution

$$k_{\max}^{\mathrm{pref}} \sim N^{1/2}. \tag{10}$$

Though fluctuations are large, Eq. (10) precisely captures the trend of the empirical data in Fig. 3(b).

3. Degree–degree correlations

Clearly the degree distribution is not an exhaustive characterization of the directory trees. It is quite evident that two networks can display the same degree distribution but show different correlation patterns in the wiring. Going one step beyond now, we ask if on average the neighbors of highly connected nodes tend to have high degrees or low degrees. Such *degree–degree correlations* are captured by the *assortativeness* [22,23] defined as

$$r = \frac{\langle kl\rangle - \langle k\rangle^2}{\langle k^2\rangle - \langle k\rangle^2}, \tag{11}$$

where the averaging $\langle\rangle$ is taken over all *links* in the tree, such that the quantity r is the Pearson correlation between the degree k at one end of a link and the degree l at the other. The degree correlations are categorized into assortative ($r > 0$), neutral ($r = 0$) and disassortative ($r < 0$). Technological networks, such as the WWW and the Internet, have been found to be disassortative, i.e. nodes of large degree have a significantly increased fraction of neighbors of low degree [22–24].

Fig. 4 shows the (negative) correlation coefficients $-r$ of the directory trees for different users. All trees we have analyzed are disassortative ($r < 0$) in agreement with previous results on artificial networks. However, the magnitude of r decreases as N grows. The trees generated by the preferential attachment model show qualitatively the same behavior. Moreover, $|r|$ decays algebraically with N in the model. A power law fit $r \propto N^{-\alpha}$ results in $\alpha = 0.35 \pm 0.04$ for the data points of the directory trees and $\alpha^{\mathrm{pref}} = 0.350 \pm 0.001$ for trees from the preferential attachment model. However, for given system size the disassortative mixing in the directory trees is by a factor 1.3 larger than in the model.

4. Distances

A natural way of characterizing the shape of an object is to compare its volume with its length. For instance, for a growing object, a dimension can be defined by observing the scaling of the length with the volume. The volume of a directory tree is the number of nodes N. A measure of length on a tree is based on the chemical distance dist_{ij} defined as the number of links contained in a shortest path between nodes i and j. Note that in a tree the path between any two nodes is unique so the minimality in the definition is not required here. We consider the sum of distances from the root with the index $i = 1$

$$\Lambda = \sum_{j=1}^{N} \mathrm{dist}_{j1}, \tag{12}$$

which is easy to estimate for the growth models. For both homogeneous and preferential attachment the evolution of Λ follows

$$\Lambda(N+1) = \Lambda(N) + \Lambda(N)/N + z. \tag{13}$$

A first-order approximation of the solution is $\Lambda(N) = zN\ln N$. For homogeneous attachment $z = 1$, because the randomly chosen parent node's expected distance is Λ/N and the new node is one step further from the root than the parent node is. For preferential attachment $z = 1/2$. This results directly from an equivalent formulation of the preferential attachment rule: choose a node at random and then attach the new node either to the chosen node itself or to its parent node with probability $1/2$. The average distance from the root $\lambda = \Lambda/N$ is

$$\lambda^{\mathrm{hom}}(N) = \ln N \tag{14}$$

$$\lambda^{\mathrm{pref}}(N) = \frac{1}{2}\ln N, \tag{15}$$

for homogeneous and preferential attachment, respectively.

Fig. 5(a) shows for the 63 directory trees that average distances λ from the root increase as the logarithm of system size. The best fit of the form $\lambda(N) = z\ln N + b$ yields $z = 0.54$ and $b = 0.02$. The preferential attachment model ($z = 1/2$) reproduces the behavior of average distances more accurately.

Further insight into the shape of the trees is gained by considering the change of λ when the trees are rewired. The two rewiring procedures *randomization* and *compactification* both conserve the degree distribution. For a rewiring step we first divide the tree by breaking a randomly chosen link (ij). In the component containing the root we break a randomly chosen link (lm). The three components are then reassembled by establishing links (il) and (jm) [25]. A randomized tree is obtained after $10N$ random rewiring steps. In the procedure of compactification a rewiring step is accepted only if it does not increase the average distance λ. Otherwise we undo the step. Rewiring is iterated until λ cannot be reduced by further steps.

In Fig. 5(b) average distances after rewiring are plotted against the distances on the original trees. Apart from small

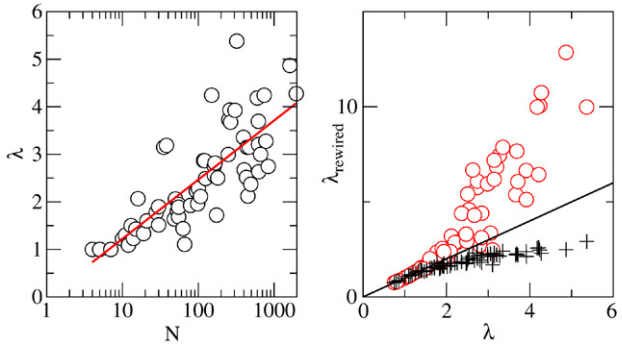

Fig. 5. Left: average distance λ of nodes from the root of the directory tree as a function of tree size N. The solid line is the result of a logarithmic fit. Right: average distance λ_{rewired} from the root after rewiring. The two rewiring modes are randomization (○) and compactification (+), see main text for details. The straight line follows $\lambda_{\text{rewired}} = \lambda$, that is invariant under rewiring. Each value λ_{rewired} for randomization is an average over 1000 randomized trees, each obtained after $10N$ random rewiring steps.

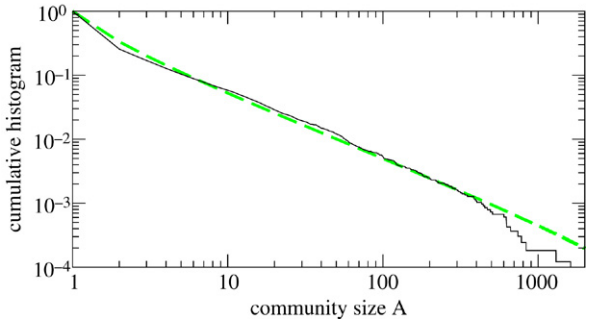

Fig. 6. Size distribution of the 16,452 communities in the 63 directory trees (solid curve). The size distribution from the preferential attachment model is also shown for comparison (dashed curve). The latter distribution has been averaged over 100 independently generated trees of size $N = 16{,}452$.

trees (original $\lambda < 2$), the rewiring significantly influences the average distances. By compactification $\lambda > 2$ is typically reduced to values $\lambda_{\text{rewired}} \approx 2$. Randomization approximately doubles the average distance on the larger trees. With respect to distance, the structure of the original directory trees is smaller than their random counterpart but not optimal, as the distance to the root can be reduced by compactification.

5. Community structure

Another distinctive property of many networks is its *community* or *branch structure*. An intuitive definition of a community is a subset of nodes which have most of their links between the members of the community and a few of them with members outside the community. Communities in a social network could represent a social group while a community on the WWW might represent pages on related topics. It is thus evident that being able to detect the communities in a network can help us to uncover its topology and make more efficient use of the networks. For instance, in the case of the WWW it can help us to locate information in fewer search steps.

A *community* is here defined as the set of directories and files contained recursively in a directory. Given a node i, the size of its community is represented by A_i. In Fig. 6 the distribution of community sizes follows a power law $P(A) \sim A^{-\tau}$, with an exponent $\tau = 2.0$. We have checked that the scaling is present not only for the aggregate statistics shown here but also holds for the individual trees. In the case of a tree, the distribution of A indicates the probability that deleting a randomly selected directory would remove a fraction of files. For comparison, we have also plotted the community distribution for trees of the same size generated with the preferential attachment algorithm, for which the scaling exponent $\tau = 2.0$ has been reported [26].

It is interesting to note that a similar scaling for the community of scientists [27,28] has been recently reported. There the network is constructed linking scientists that have authored a paper together. The exponent of the distribution of communities follows a power-law behavior with an exponent $\tau \simeq 2$. The Internet also seems to belong to the same class [29]. However, a different class is formed by river networks [8,30,31], informal networks in organizations [12] and jazz musician networks [28] where the corresponding exponent gives a value ∼1.45.

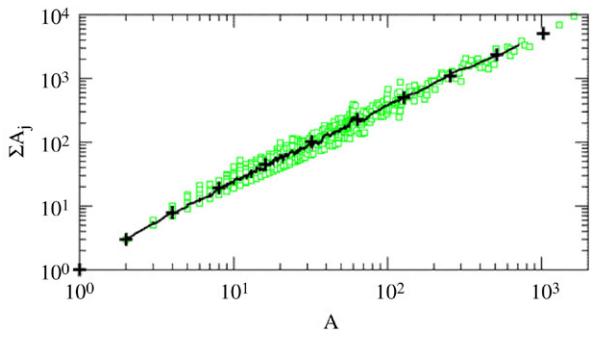

Fig. 7. Allometric scaling. The directory trees yield a total of 16,452 allometric data points (squares). The solid curve is a running average with window size 20. For comparison the values from the preferential attachment model are shown (+).

5.1. Allometric scaling

A more sophisticated characterization of the community structure of a tree can be given in terms of what is called allometric scaling. Given a directory tree we first calculate for a directory i its community size A_i. Then, we also compute the sum $C_i = \sum_{j \in V(i)} A_j$ where $V(i)$ runs over all directories in i including itself. It has been found that the dependence of C on A is a power law $A^{-\eta}$ with an exponent η that is universal in river networks 3/2 and vascular networks in animals 4/3. It has been proven [7,9,32] that in an *efficient* network the exponent η is related to the embedding dimension D as $\eta = (D+1)/D$. This result predicts for river networks ($D = 2$) $\eta = 3/2$, while for the vascular system ($D = 3$) $\eta = 4/3$, in agreement with the empirical results. However, very recently it has been reported that in plants the scaling exponent is 1 [33], challenging the previous picture. This analysis has also been performed in food webs. Food webs, with an exponent around 1.13, seem to deviate slightly from the efficiency hypothesis that predicts an exponent 1 ($D \to \infty$), possibly due to the competition between species [13]. In the case of directory trees

we first observe that it follows a scaling law with an estimated exponent $\eta = 1.2$, see Fig. 7. This exponent also deviates from the efficiency hypothesis. It is interesting to note that the preferential model shows an excellent agreement with the empirical data. The deviations from the efficiency hypothesis can be explained in terms of logarithmic corrections to the scaling law. The allometric scaling can be rewritten as $C_i = \sum_{j\in V(i)} \mathrm{dist}_{ji} = A_i(1+\lambda_i)$, where λ_i is the average distance of the nodes in a community to node i. Taking into account Eqs. (14) and (15) we obtain the scaling relation $C \sim A \ln A$. Thus, the deviation from the linear scaling could be explained by means of a logarithmic correction. It would be interesting to test whether this result explains the deviations observed in food webs [34,35].

6. Conclusions

Nowadays much information is stored electronically and available on the WWW. Understanding how information is stored by computer users is very relevant, for instance, for the design of search engines. We have precisely characterized the topological features of directory and file trees generated by computer users in a research institution. Having access to many different trees we are able to capture general features shared by many users, but at the same time observe the differences. This approach complements other studies which concentrate on complex networks as the outcome of a collective phenomena.

We have observed that independent users generate trees with similar structural properties. The main topological features of the trees are

(1) a broad degree distribution;
(2) the average distance to the root increases logarithmically with system size;
(3) negative degree–degree correlations that decay with the size of the tree with a power law;
(4) distribution of community or branch size follows a power law with an exponent $\tau = 2$;
(5) allometric scaling with an exponent close to 1.

We have shown that these general properties can be captured with a model of growing trees with preferential attachment, but not with a model with random attachment. The model correctly predicts the scaling of the distance to the root with tree size, the distribution of community sizes and predicts a linear growth of the allometric scaling with logarithmic corrections.

Despite these strong similarities between trees from the data set and those from the model, one cannot conclude that the preferential attachment model is a complete description of how users build trees. A crucial assumption we have made is that trees are generated mainly by iterative addition of nodes. In reality, this means that users generate new directories much more frequently than they delete and relocate existing ones. When this holds and the growth generates a scale-free degree distribution, as is the case here, the attachment probability must be asymptotically linear in the degree. Preferential attachment is a sufficient and *necessary* condition for a scale-free degree distribution in growing networks [15,36]. Thus the next step of research is to record trees from the same users repeatedly, e.g. on a daily basis. This will allow us to check the aforementioned assumption that iterative growth is the dominant process and other operations can be neglected.

With such time-resolved data we expect to obtain insight into users' behaviour. For instance, spatiotemporal correlations between attachment events are likely because users tend to attach new directories to the same subtree (community) as long as their work sticks to one given topic. Insight into the mechanisms on the behavioral level should lead to a more refined model. Ideally such a model would be mechanistic in that it explains the users' attachment choices in terms of the features of the filed content.

The question whether optimization is a driving force of the tree formation deserves further effort. In terms of observed exponents and proposed models, directory trees resemble structures of systems known to be under optimization pressure (see the introduction of this paper). Going beyond these analogies, optimality can be formulated rigorously in terms of the cost of navigation and search of a given network structure [37–39]. Search performance, however, crucially depends on previous knowledge about the network [40]. Inevitably, the user's partial memory of her/his own file locations must enter an appropriate model of directory trees in terms of optimization. However, it is unlikely that directory trees are optimal structures with respect to a universal non-changing measure of cost. If users tried to keep their trees optimal under the influx of new content they would have frequently to discard the structure and design it from scratch. In contrast to textbooks, websites and other *designed* structures, directory trees are *tinkered* [3].

In summary, by describing and modeling directory trees generated by single users we advance in our understanding on how information is stored and managed. Our approach complements other studies that focus, for example, on the WWW that emerges from the interaction of many users. We have found that similar principles (preferential attachment) explain the main features of the structures.

Acknowledgments

We acknowledge financial support from MCyT (Spain) through project CONOCE2, from Dutsche Forschungsgemeinschaft through Bioinformatics Initiative BIZ-6/1-2 and from Deutscher Akademischer Austauschdienst (DAAD).

References

[1] R.M. Shiffrin, K. Borner, Proc. Natl Acad. Sci. USA 101 (2004) 5183–5184.
[2] S. Lawrence, C.L. Giles, Accessibility of information on the web, Nature 400 (1999) 107–109.
[3] F. Jacob, Science 196 (1997) 1161.
[4] R.V. Sole, R. Ferrer, J.M. Montoya, S. Valverde, Complexity 8 (2002) 20.
[5] S. Valverde, R. Ferrer-Cancho, R.V. Sole, Europhys. Lett. 60 (2002) 512.
[6] C.R. Myers, Phys. Rev. E 68 (2003) 046116.
[7] J.R. Banavar, A. Maritan, A. Rinaldo, Size and form in efficient transportation networks, Nature 399 (1999) 130–132.
[8] I. Rodríguez-Iturbe, A. Rinaldo, Fractal River Basins: Chance and Self-Organization, Cambridge Univ. Press, New York, 1996.
[9] G.B. West, J.H. Brown, B.J. Enquist, A general model for the origin of

allometric scaling laws in biology, Science 276 (1997) 122–126.
[10] G.B. West, J.H. Brown, B.J. Enquist, The fourth dimension of life: Fractal dimension and allometric scaling of organisms, Science 284 (1999) 1677–1679.
[11] M. Girvan, M.E.J. Newman, Proc. Natl Acad. Sci. USA 99 (2002) 7821.
[12] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, A. Arenas, Self-similar community structure in organisations, Phys. Rev. E 68 (2003) 065103(R).
[13] D. Garlaschelli, G. Caldarelli, L. Pietronero, Nature 423 (2002) 165–168.
[14] R. Ferrer i Cancho, R.V. Solé, Optimization in complex networks, in: Statistical Physics of Complex Networks, in: Lecture Notes in Physics, Springer, Berlin, 2004.
[15] K.A. Eriksen, M. Hörnquist, Phys. Rev. E 65 (2002) 017102.
[16] K. Klemm, V.M. Eguíluz, M. San Miguel, Scaling in the structure of directory trees in a computer cluster, Phys. Rev. Lett. 95 (2005) 128701.
[17] S.H. Strogatz, Exploring complex networks, Nature 410 (2001) 268–276.
[18] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Rev. Modern Phys. 74 (2002) 47–97.
[19] S.N. Dorogovtsev, J.F.F. Mendes, Adv. Phys. 51 (2002) 1079–1187.
[20] M.E.J. Newman, The structure and function of complex networks, SIAM Rev. 45 (2003) 167–265.
[21] A.L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509–512.
[22] M.E.J. Newman, Assortative mixing in networks, Phys. Rev. Lett. 89 (2002) 208701.
[23] M.E.J. Newman, Mixing patterns in networks, Phys. Rev. E 67 (2003) 026126.
[24] J. Park, M.E.J. Newman, Origin of degree correlations in the Internet and other networks, Phys. Rev. E 68 (2003) 026112.
[25] S. Maslov, K. Sneppen, Specificity and stability in topology of protein networks, Science 296 (2002) 910–913.
[26] P.L. Krapivsky, S. Redner, Organization of growing random networks, Phys. Rev. E 63 (2001) 066123.
[27] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, D. Parisi, Proc. Natl Acad. Sci. USA 101 (2004) 2658–2663.
[28] A. Arenas, L. Danon, A. Díaz-Guilera, P. Gleiser, R. Guimera, Community analysis in social networks, European J. Phys. B 38 (2004) 373.
[29] G. Caldarelli, R. Marchetti, L. Pietronero, The fractal properties of internet, Europhys. Lett. 52 (2000) 386.
[30] A. Rinaldo, I. Rodriguez-Iturbe, R. Rigon, E. Ijjasz-Vazquez, R.L. Bras, Self-organized fractal river networks, Phys. Rev. Lett. 70 (1993) 822–825.
[31] A. Maritan, A. Rinaldo, R. Rigon, A. Giacometti, I. Rodriguez-Iturbe, Scaling laws for river networks, Phys. Rev. E 53 (1996) 1510–1515.
[32] J.R. Banavar, J. Damuth, A. Maritan, A. Rinaldo, Supply-demand balance and metabolic scaling, Proc. Natl Acad. Sci. USA 99 (2002) 10506–10509.
[33] P.B. Reich, M.G. Tjoelker, J.-L. Machado, J. Oleksyn, Universal scaling of respiratory metabolism, size and nitrogen in plants, Nature 439 (2006) 457.
[34] J. Camacho, A. Arenas, Food-web topology: Universal scaling in food-web structure? Nature 435 (2005) E3.
[35] D. Garlaschelli, G. Caldarelli, L. Pietronero, Food-web topology: Universal scaling in food-web structure? Nature 435 (2005) E4 (reply).
[36] This implication appears to disagree with the users' own perception. When confronted with the results and analysis of the present paper, users found iterative growth plausible as the dominating process for building directory trees. The preferential attachment rule, however, does not agree with the way users think they place new directories.
[37] J.M. Kleinberg, Navigation in a small-world, Nature 406 (2000) 845.
[38] X. Zhu, J. Yu, J.C. Doyle, Proc. IEEE Infocom, 2001, pp. 1617–1626.
[39] M.R. Roberson, D. ben-Avraham, Kleinberg navigation in fractal small-world networks, Phys. Rev. E 74 (2006) 017101.
[40] M. Rosvall, P. Minnhagen, K. Sneppen, Navigating networks with limited information, Phys. Rev. E 71 (2005) 066111.