

## A local hierarchy theory for acyclic digraphs

Peer-reviewed author version

EGGHE, Leo & ROUSSEAU, Ronald (2004) A local hierarchy theory for acyclic digraphs. In: *Mathematical and Computer Modelling*, 39(1). p. 107-117.

DOI: [10.1016/S0895-7177\(04\)90510-9](https://doi.org/10.1016/S0895-7177(04)90510-9)

Handle: <http://hdl.handle.net/1942/759>

# A LOCAL HIERARCHY THEORY FOR ACYCLIC DIGRAPHS

by

L. Egghe, LUC, Universitaire Campus, B-3590 Diepenbeek, Belgium<sup>1</sup>

and

UA, Universiteitsplein 1, B-2610 Wilrijk, Belgium

and

R. Rousseau, LUC, Universitaire Campus, B-3590 Diepenbeek, Belgium

and

UA, Universiteitsplein 1, B-2610 Wilrijk, Belgium

---

**Keywords :** Local hierarchy theory, global hierarchy theory, acyclic digraphs, generalized Lorenz curve, citation analysis, organizational charts

## **ABSTRACT**

Local hierarchy theory focuses on direct links in acyclic digraphs. In- and out-degrees are used to determine the local hierarchical number for each vertex in the graph. Together, these local hierarchical numbers form a vector through which hierarchical properties are studied. The main tool, leading to a partial order of acyclic digraphs is a form of generalized Lorenz curve. Gini-like measures respecting this partial order can be derived. Local hierarchy theory is then the theory related to this particular partial order. Results have possible applications in administration and business organizational charts and in citation analysis. In the latter a direct link represents a reference or a citation of a document. Finally, we study rooted trees as a concrete example of local hierarchy theory.

---

<sup>1</sup> Permanent address.

## I. Introduction

In this article we complement the global hierarchy theory (in short : GHT) (studied in Egghe, 2001) by a local theory, referred to as local hierarchy theory (in short : LHT). Global as well as local hierarchy theory can be used in studies of citation networks, business organization charts, trees, and many other networks. As we assume that the underlying graphs must be acyclic and directed (see further for a precise definition) this theory can, however, not be applied to web hyperlinks or collaboration networks, as these are either undirected or contains loops (cases where web page W is linked to web page P, while also web page P is linked to W). We next recall some basic definitions and results from general graph theory.

A directed graph (in short digraph)  $G(V,E)$  consists of a set  $V = \{1, \dots, N\}$  of vertices or nodes, and a set  $E$  of ordered pairs of the form  $(i,j)$  where  $i$  and  $j$  are in  $V$ . An ordered pair  $(i,j)$  is called an edge (or more precisely : a directed edge). The set of edges of a given graph  $G$  is denoted as  $E(G)$ . Node  $i$  is called the initial node and node  $j$  is called the terminal node of the edge  $(i,j)$ . A directed path, or chain, from node  $i$  to node  $j$  is a set of edges  $(v_k)$ , such that the terminal node of edge  $v_k$  coincides with the initial node of edge  $v_{k+1}$  and such that node  $i$  is the initial node of edge  $v_1$ , and node  $j$  is the terminal node of edge  $v_N$ . If node  $i$  coincides with node  $j$  the directed path is a directed circuit (or loop). A directed graph is called acyclic, or loopless, if it contains no directed circuits. A directed graph is weakly connected if there exists a path between any two nodes in the underlying undirected graph (Wilson, 1972). We will always assume this to be true. If this is not the case then the theory can be applied to weakly connected components of the graph.

Definition : in- and out-degree (Chen, 1971)

For a directed graph  $G$ , the number  $\alpha_j^+$  of edges of  $G$  having node  $j$  as their initial node, is called the out-degree of node  $j$ . Similarly, the number  $\alpha_j^-$  of edges in  $G$  having node  $j$  as their terminal node is called the in-degree of node  $j$ .

We put

$$\alpha_j = \alpha_j^+ - \alpha_j^- \quad (1)$$

This parameter  $\alpha_j$  characterizes the flow through node  $j$ . Clearly  $\alpha_j^+ + \alpha_j^-$  is equal to the number of edges of  $G$  incident with node  $j$ .

Since every edge is outgoing from a node and terminating at another, it is evident that the number  $\varepsilon$  of edges of  $G$  is related to the degrees of its nodes by the following equation (Chen, 1971, p.29) :

$$\varepsilon = \sum_j \alpha_j^+ = \sum_j \alpha_j^- \quad (2)$$

$$\text{or} \quad \sum_j \alpha_j = 0 \quad (3)$$

where the summation is over all nodes of graph  $G$ . Equation (3) implies that the Lorenz theory for vectors consisting of coordinates summing to zero can be applied on the vector  $X = (\alpha_j)_{j=1, \dots, N}$  (Egghe, 2001). Such vectors will be referred to as zero-sum vectors. In (Egghe, 2001) the author studied acyclic digraphs from a global point of view. Results reflected the overall inequality among bosses and subordinates in the network, using lengths of all possible paths between vertices. This led to global hierarchy theory (GHT). In the present paper only the numbers of immediate superiors and immediate subordinates define the inequality of the network. For this reason we refer to this approach as local hierarchy theory (LHT).

Egghe (2001) showed that in an  $N$ -node network vectors  $X$  yielding maximal and minimal Lorenz curves for the GHT are given by :

$$X = (x, \underbrace{0, \dots, 0}_{N-2}, -x) \quad (4)$$

where  $x > 0$  for the maximal ones and

$$X = (\underbrace{x_1, \dots, x_i}_i, \underbrace{-y_1, \dots, -y_{N-i}}_{N-i}) \quad (5)$$

$x > 0, y > 0$  for the minimal ones.

In LHT we form vectors of the form  $X = (\alpha_1, \alpha_2, \dots, \alpha_N)$ ,  $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_N$ . This implies that (in an organizational chart) the first coordinates correspond to people having many immediate subordinates, while the last ones are people who are immediate subordinate to many others. In a citation network where a link means "is cited by" the first alphas represent articles (or authors) that received many citations. However, in a citation network where a link means 'cites' the most-cited articles or authors are those represented by the smallest alphas.

As for GHT (Egghe, 2001) we will in this paper investigate for LHT which graphs yield maximal and minimal Lorenz curves (Section II). In Section III we will give a direct relation between GHT and LHT. Finally, in Section IV we study rooted trees as a concrete example of local hierarchy theory.

The present study, as well as the one in (Egghe, 2001) is related to, but different, from the hierarchy theory developed by Botafogo, Rivlin and Shneiderman (1992). These authors used a global approach, but base their theory on shortest distances between nodes. In this way they consider only a part of the structure present in the network. Their approach was further adapted to citation networks by De Bra (2000). In these two articles the term stratum is used for a metric indicating how deep or linear a link structure is. We think, however, that our approach, using a revised form of Lorenz-curves, and measures derived from these curves, is more precise.

## **II Local hierarchy theory (LHT) derived from generalized Lorenz curves for zero-sum vectors**

### **II.1. Lorenz curves for zero-sum vectors $X = (x_1, \dots, x_N)$**

In this section we recall the theory developed by Egghe (2001) for zero-sum vectors  $X=(x_1, \dots, x_N)$ .

We assume that not all  $x_i$  are zero and that the  $x_i$  are decreasing. Denote

$$I_+ = \{i \in \{1, \dots, N\} \mid x_i > 0\} \quad (6)$$

$$I_- = \{i \in \{1, \dots, N\} \mid x_i < 0\} \quad (7)$$

Hence

$$\sum_{k=1}^N x_k = \sum_{i \in I_+} x_i + \sum_{i \in I_-} x_i = 0 \quad (8)$$

so

$$\sum_- = : \sum_{i \in I_-} x_i = - \sum_{i \in I_+} x_i \quad (9)$$

Equation (9) enables us to develop a concentration theory (or inequality theory) for vectors  $X$  for which the coordinates add up to zero, hereby studying the concentration in  $(x_i)_{i \in I_+}$  as well as the concentration in  $(x_i)_{i \in I_-}$ . We proceed as follows. Calculate

$$a_i = \frac{x_i}{\sum_-} \quad (10)$$

for all  $i=1, \dots, N$ . Because of (8) and (9) we have

$$\sum_{i \in I_+} a_i = 1 \quad (11)$$

$$\sum_{k=1}^N a_k = 0 \quad (12)$$

We now form the polygonal curve connecting  $(0,0)$  with  $\left(\frac{1}{N}, a_1\right)$ ,  $\left(\frac{1}{N}, a_1\right)$  with  $\left(\frac{2}{N}, a_1+a_2\right)$  and so on until we reach  $(x,1)$ , where

$$x = \frac{|I_+|}{N}. \quad (13)$$

Then we connect  $(x,1)$  to  $(y,1)$ , where

$$y = \frac{N-|I_-|}{N}. \quad (14)$$

Finally, we connect  $(y,1)$  to  $(1,0)$  via the points  $\left(\frac{i}{N}, \sum_{k=1}^i a_k\right)$  where  $i \in I_-$ . Note that  $x \leq y$  since  $|I_-| + |I_+| \leq N$ . If no  $x_i$  is zero, then  $x=y$  since then  $|I_-| + |I_+| = N$  then. An example of this kind of generalized Lorenz curve is given in Fig. 1.

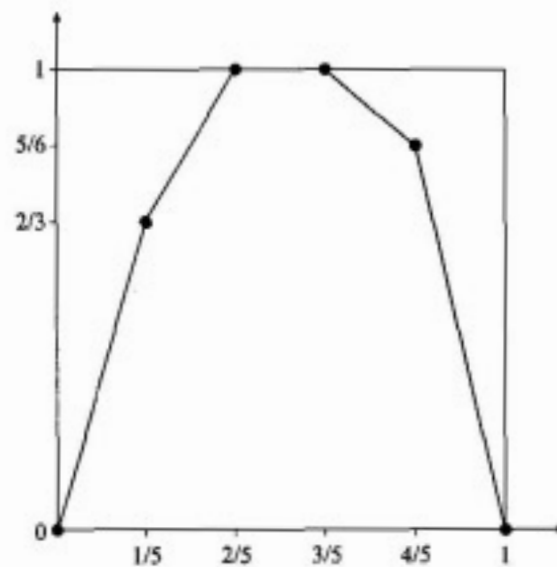


Fig. 1 Lorenz curve  $L_X$  for  $X=(4,2,0,-1,-5)$  connecting  $(0,0), (\frac{1}{5}, \frac{2}{3}), (\frac{2}{5}, 1), (\frac{3}{5}, 1), (\frac{4}{5}, \frac{5}{6}), (1,0)$ .

Intuitively speaking,  $L_X$  consists of a "Lorenz curve" for the  $(x_i)_{i \in I_+}$  (from  $(0,0)$  to  $(x,1)$ ) and of a "Lorenz curve" for the  $(x_i)_{i \in I_-}$  (from  $(y,1)$  to  $(1,0)$  and mirrored over the vertical line with abscissa  $y$ ). This is why we have here a method of measuring the concentration in

the  $(x_i)_{i \in I_+}$  as well as in the  $(x_i)_{i \in I_-}$ . The "total" degree of inequality can then be compared with that of another vector as follows.

Let, indeed, be  $X=(x_1, \dots, x_N)$  and  $X'=(x'_1, \dots, x'_N)$  be two decreasing vectors such that

$$\sum_{k=1}^N x_k = \sum_{k=1}^N x'_k = 0 \quad (15)$$

We say that  $X'$  is larger than  $X$  in the Lorenz sense, denoted  $X \prec X'$  if  $L_X \leq L_{X'}$ . If  $X \neq X'$ , then  $X'$  represents a more concentrated situation in both the positive and negative values. This will enable us, in applications (see Egghe (2001) and further in this paper) to measure the hierarchical degree (both in domination and subordination as one system) in a digraph, with obvious practical applications.

## II.2. LHT and graphs yielding the maximal and minimal Lorenz curves.

### II.2.1. LHT

The Lorenz theory explained above is applied to the vector  $X=(\alpha_1, \dots, \alpha_N)$  where  $\alpha_i$  is as in section I.1. If  $X'=(\alpha'_1, \dots, \alpha'_N)$  is a second vector (derived from another graph, i.e. another hierarchical situation), the relation  $X \prec X'$ , introduced in the previous subsection, meaning  $L_X \leq L_{X'}$  (the Lorenz curve of  $X$  is below the one of  $X'$ ), expresses the local hierarchical degree of the two graphs, in the sense that the  $X'$ -situation has a higher local hierarchical degree (i.e. inequality) than the  $X$ -situation. Concrete good measures of local hierarchical degree can be given, e.g. (see Egghe (2001))

$$\sum_{i=1}^N \left( \frac{\alpha_i}{\sum \cdot} \right)^2 \quad (16)$$

or, simply, the area under the Lorenz curve, i.e. the area between  $L_X$  and the x-axis.

LHT has the following properties :

- (i) The higher the inequality between the direct bosses ( $\alpha_i > 0$ ), the higher  $L_X$ , hence the higher the local hierarchical degree ;



- (ii) The higher the inequality between the direct subordinates ( $\alpha_i < 0$ ), the higher  $L_x$ , hence, again, the higher the local hierarchical degree.

These properties are direct consequences from the construction of the Lorenz curve.

We will now study maximal and minimal Lorenz curves, and find which graphs (hierarchies) yield these maximal and minimal curves. These graphs have high respectively low local hierarchical degrees.

### II.2.2. Graphs yielding the maximal Lorenz curve in LHT.

Recall (from Egghe (2001) or (4)) that the maximal Lorenz curve  $L_x$  is obtained for

$$X = (x, \underbrace{0, \dots, 0}_{N-2}, -x),$$

where  $x > 0$ .

Which graphs yield such a maximal situation in the LHT ?

Definition II.2.2.1 : Let  $G$  be a weakly connected digraph without loops. We say that  $G$  is a C-string if  $G$  is the union of chains all starting in the same point (called 1) and ending in the same point (called  $N$ ) and which do not intersect elsewhere except (possibly) in vertices which are not directly linked.

### Examples II.2.2.2 :

1. All graphs yielding the maximal Lorenz curves in the GHT (Egghe (2001)) - see Fig. 2.

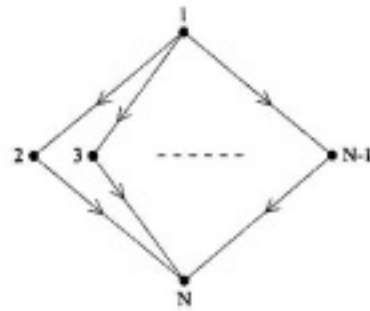


Fig. 2 Graph yielding the largest hierarchical Lorenz curve in the GHT

2. All chains - see Fig. 3.



Fig. 3 A general chain of N vertices

3. Combinations of the above Figs. - see Fig. 4.



Fig. 4a



Fig. 4b

Fig. 4 Two generalizations of Figs. 2-3, being also C-strings.

4. Note that the graph in Fig. 5 is not a C-string.



Fig. 5 Example of a graph that is not a C-string

We have the following characterization of graphs  $G$  yielding  $L_{\max}$  (i.e. the maximal Lorenz curve) in LHT :

Theorem II.2.2.3 : A graph  $G$  yields  $L_{\max}$  in LHT iff  $G$  is a C-string.

Proof : If  $G$  is a C-string it is clear that, in LHT, this graph yields  $L_{\max}$  since for all vertices  $i \in \{2, \dots, N-1\}$  the number of in-links is equal to the number of out-links, hence  $\alpha_i = 0$ . It is also clear that  $\alpha_1 = -\alpha_N$ .

Conversely, let  $G$  be a weakly connected digraph without loops such that

$$(\alpha_1, \dots, \alpha_N) = (x, \underbrace{0, \dots, 0}_{N-2}, -x), \quad x > 0.$$

(i) There exists in G a chain between 1 and N.

Indeed, since  $\alpha_1 = x > 0$ , there is at least one direct link starting in 1. If such a direct link arrives in  $i \in \{2, \dots, N-1\}$  there must be another direct link starting in  $i$  since  $\alpha_i = 0$ . This can only stop in N.

(ii) Node 1 cannot have a direct in-link.

Suppose 1 has a direct in-link. Let  $i \in \{2, \dots, N\}$  be such that  $(i, 1) \in E$ . If  $i = N$ , then we have the loop  $N \rightarrow 1 \rightarrow \dots \rightarrow N$ , by (i) which is excluded. So  $i \in \{2, \dots, N-1\}$ , hence  $\alpha_i = 0$ . So there is a vertex  $j \in \{1, \dots, N\} \setminus \{i\}$  such that  $(j, i) \in E$ . If  $j = 1$  we have the loop  $1 \rightarrow i \rightarrow 1$  which is excluded. If  $j = N$  we have the loop  $N \rightarrow i \rightarrow 1 \rightarrow \dots \rightarrow N$ , which is also excluded. So  $j \in \{2, \dots, N-1\} \setminus \{i\}$  and hence  $\alpha_j = 0$ . The argument is repeated until a last  $k \in \{2, \dots, N-1\}$  such that  $k \rightarrow \dots \rightarrow j \rightarrow i \rightarrow 1$ . But then  $\alpha_k \neq 0$  which is false. So 1 has no direct in-link.

Exactly the same argument proves that

(iii) Node N cannot have a direct out-link.

In conclusion, since

$$(\alpha_1, \dots, \alpha_N) = (x, \underbrace{0, \dots, 0}_{N-2}, -x),$$

in 1 depart exactly  $x$  direct links and in N arrive exactly  $x$  direct links. Further, in each of the vertices  $i \in \{2, \dots, N-1\}$ , whenever an in-link arrives, also an out-link departs, since  $\alpha_i = 0$ . So, we have a finite number of chains between 1 and N. They can intersect in vertices  $i \in \{2, \dots, N-1\}$  but, if  $i, j$  are two of such intersections,  $(i, j)$  nor  $(j, i)$  can belong to  $E$  since in that case  $\alpha$  of the first vertex would be positive and  $\alpha$  of the second vertex would be negative which is not possible since all  $\alpha$ -values are zero for vertices  $i, j \in \{2, \dots, N-1\}$ .  $\square$

Corollary II.2.2.4 : If G is a graph yielding  $L_{\max}$  in GHT, it yields  $L_{\max}$  in LHT.

Proof : This follows immediately from the previous theorem and the corresponding result in Egghe (2001) - see also the first examples in II.2.2.2.  $\square$

### II.2.3. Graphs yielding a minimal Lorenz curve in LHT

We were not able to find a full characterization of graphs yielding a minimal Lorenz curve in LHT. It is, however, clear that there is a wide variety of them. It can also be shown that all graphs yielding a minimal Lorenz curve in GHT also yield a minimal Lorenz curve in LHT.

Theorem II.2.3.1 : If  $G$  is a graph yielding a minimal Lorenz curve in GHT, it yields a minimal Lorenz curve in LHT.

Proof : The characterization of graphs, yielding a minimal Lorenz curve in GHT, given in Egghe (2001) is as follows (interpreted in the terminology of this paper) : upon a permutation of  $\{1, \dots, N\}$ , we have that only the vertices  $1, \dots, i$  ( $i \in \{1, \dots, N\}$ , a free parameter) have an equal number of direct out-links to the vertices  $i+1, \dots, N$  which have an equal number of direct in-links and no links between the vertices  $1, \dots, i$  or between the vertices  $i+1, \dots, N$  exist. Since in this graph there are only direct links, LHT is the same as GHT, hence it yields a minimal Lorenz curve in LHT.  $\square$

It is also clear that the set of graphs yielding a minimal Lorenz curve in LHT is a strict superset of the set of graphs yielding a minimal Lorenz curve in GHT. The next examples illustrate this : the following graphs yield minimal Lorenz curves in LHT.

#### Examples II.2.3.2.

1. For the graph in Fig.6a we have  $\alpha_1 = \alpha_2 = 1$ ,  $\alpha_3 = \alpha_4 = -1$ , hence  $X = (1, 1, -1, -1)$ , yielding a minimal Lorenz curve, by (5). This graph is not of the type described in Theorem II.2.3.1, since  $(3, 4) \in E$ .

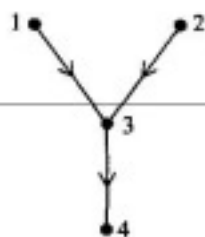


Fig. 6a

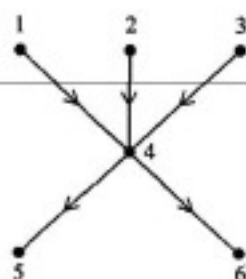


Fig. 6b

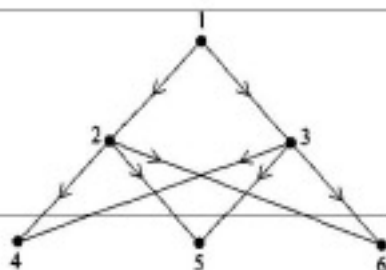


Fig. 6c

Fig. 6 Graphs yielding a minimal Lorenz curve in LHT but not in GHT.

2. For the graph in Fig. 6b we have  $\alpha_1 = \alpha_2 = \alpha_3 = 1$ ,  $\alpha_4 = \alpha_5 = \alpha_6 = -1$ . Note that  $(4,5), (4,6) \in E$  so that this graph is not of the type described in Theorem II.2.3.1.
3. The graph in Fig. 6c has the  $\alpha$ -values :  $\alpha_1 = \alpha_2 = \alpha_3 = 2$ ,  $\alpha_4 = \alpha_5 = \alpha_6 = -2$ . Further  $(1,2), (1,3) \in E$  and hence also this graph is not of the type described in Theorem II.2.3.1.

### III. Relations between GHT and LHT.

Instead of working with the vector  $X=(\alpha_1,\dots,\alpha_N)$  (as defined in the introduction) in LHT, GHT uses the vector  $X=(\sigma_1,\dots,\sigma_N)$  where, for each  $i=1,\dots,N$

$$\sigma_i = \sigma_i^+ - \sigma_i^- , \quad (17)$$

with

$$\sigma_i^+ = \text{the sum of the lengths of all chains that start in } i \quad (18)$$

and

$$\sigma_i^- = \text{the sum of the lengths of all chains that end in } i \quad (19)$$

(Egghe (2001)). Since all chains are considered, GHT is indeed a global hierarchy theory and is different from models where one only considers the distance between two vertices (being the length of the shortest chain between them). Such a distance was e.g. used in Botafogo, Rivlin and Shneiderman (1992). The problem in GHT is the calculation of (17) which is complicated as opposed to the calculation of  $\alpha_i$ , since there only direct links are used. In this section, however, we determine a relationship between GHT and LHT in the following sense : for each  $i=1,\dots,N$ ,  $\sigma_i$  is expressed as the sum of  $\alpha_j$  and (as a recursion) a formula in which only  $\sigma_j$ , for  $j$  that have direct links with  $i$ , are appearing. We have the following theorem.

Theorem III.1 : for all  $i=1,\dots,N$

Let  $D_i^+ = \{j \in \{1,\dots,N\} | (i,j) \in E\}$

$$R_i^+ = \{j \in \{1,\dots,N\} | \text{there is a chain from } i \text{ to } j\}.$$

Then

$$\sigma_i^+ = \sum_{j \in D_i^+} (\sigma_j^+ + \#R_j^+) + \alpha_i^+ \quad (20)$$

and similarly, let

$$D_i^- = \{j \in \{1, \dots, N\} \mid (j, i) \in E\}$$

$$R_i^- = \{j \in \{1, \dots, N\} \mid \text{there is a chain from } j \text{ to } i\}.$$

Then

$$\sigma_i^- = \sum_{j \in D_i^-} (\sigma_j^- + \#R_j^-) + \alpha_i^-. \quad (21)$$

Hence, for every  $i=1, \dots, N$

$$\sigma_i = \alpha_i + \sum_{j \in D_i^+} (\sigma_j^+ + \#R_j^+) - \sum_{j \in D_i^-} (\sigma_j^- + \#R_j^-) \quad (22)$$

Proof : We only prove the result for  $\sigma_i^+$ ; the one of  $\sigma_i^-$  is similar and (22) follows from  $\sigma_i = \sigma_i^+ - \sigma_i^-$ . For each  $j \in D_i^+$  we have that the length of every chain with  $j$  as unital node increases with one unit when we consider the chain starting in  $i$ , via  $j$ . Since there are  $\#R_j^+$  such paths we have that  $\sigma_i^+$  is composed of the sum of the lengths of all such paths via  $j$ , being  $\sigma_j^+ + \#R_j^+$ , added over all  $j \in D_i^+$ . Then we have to add the lengths of all (direct) paths from  $i$  to  $j$ , being  $\#D_i^+ = \alpha_i^+$ . Hence

$$\sigma_i^+ = \sum_{j \in D_i^+} (\sigma_j^+ + \#R_j^+) + \alpha_i^+. \quad \square$$

Note : Note that, since the graph has no loops, that  $R_i^+ \cap R_i^- = \emptyset$  for all  $i=1, \dots, N$ . Note also that  $R_i^+$  and  $R_i^-$  are, essentially, the tail and the head of node  $i$ , as described in (Egghe and Rousseau, 2001). The only difference is that node  $i$  belongs to its head and tail, and not to its R-sets.



#### IV. LHT for rooted trees

A rooted tree with fixed branching factor  $b \in \mathbb{N}$  is a graph in which  $b \in \mathbb{N}$  direct links depart - say from 0 (called the root) to vertices  $1, \dots, b$  - constituting the first level. If there is a second level then each of the vertices  $1, \dots, b$  is the departure point of  $b$  new vertices, constituting the second level (consisting of  $b^2$  vertices). The number of levels  $d$  is called the depth of the tree. The vertices on the last level are called leaves. Note that the set of vertices is  $\{0, \dots, N\}$  here (for ease of notation) instead of  $\{1, \dots, N\}$  as used so far. See Fig.7 for a visualization of the construction of a rooted tree ( $b=2$  and  $d=3$ ).

LHT for such trees goes as follows. Let  $b \neq 1$ . Then  $\alpha_i^+ = b$ ,  $\alpha_i^- = 1$  for all vertices that are neither root or leaf (i.e. for  $i \in \{1, \dots, N - b^d\}$ ). Hence for these  $i$ ,  $\alpha_i = b - 1$ . If  $i = 0$  (the root) we have  $\alpha_0 = \alpha_0^+ = b$ . If  $i$  is a leaf (hence  $i \in \{N - b^d + 1, \dots, N\}$ ), we have  $\alpha_i = -\alpha_i^- = -1$ .



Fig. 7 Construction of a tree with branching factor  $b=2$  and depth  $d=3$

Consequently (see(9))

$$\begin{aligned}\sum_{\cdot} &= \sum_{i=1}^{d-1} b^i(b-1)+b \\ \sum_{\cdot} &= (b-1) \frac{b^d-b}{b-1} + b = b^d\end{aligned}\quad (23)$$

Since the vector  $X = \{\alpha_1, \dots, \alpha_N\}$  for this tree is

$$\begin{array}{c} (b, b-1, \dots, b-1, -1, \dots, -1) \\ \underbrace{\hspace{1.5cm}} \quad \underbrace{\hspace{1.5cm}} \\ \frac{b^d-b}{b-1} \text{ times} \quad b^d \text{ times} \end{array}$$

we have that its normalized form (10), used for the construction of the Lorenz curve is (divide by  $\sum_{\cdot}$ ) :

$$\begin{array}{c} \left( b^{-d+1}, \underbrace{\frac{b-1}{b^d}, \dots, \frac{b-1}{b^d}}_{\frac{b^d-b}{b-1} \text{ times}}, \underbrace{-\frac{1}{b^d}, \dots, -\frac{1}{b^d}}_{b^d \text{ times}} \right) \\ \underbrace{\hspace{1.5cm}} \quad \underbrace{\hspace{1.5cm}} \\ \frac{b^d-b}{b-1} \text{ times} \quad b^d \text{ times} \end{array}\quad (24)$$

Hence the Lorenz curve looks as in Fig.8

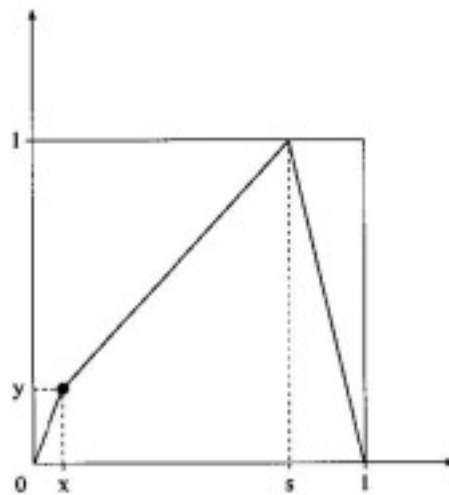


Fig. 8 Lorenz curve of a rooted tree

Here  $x = \frac{b-1}{b^{d+1}-1}$  (being  $\frac{1}{N}$ ),  $y = b^{-d+1}$  and  $s$  (being  $\frac{N-b^d}{N}$ ) is given by

$$s = \frac{b^d - 1}{b^{d+1} - 1}$$

The area under this curve (as noted before, a good measure) is

$$A = \frac{1}{2}xy + \frac{1}{2}(1-s) + y(s-x) + \frac{1}{2}(1-y)(s-x)$$

$$A = \frac{b^{2d} - 1}{2(b^{2d} - b^{d-1})} \quad (25)$$

Note that  $\lim_{d \rightarrow \infty} A = \lim_{b \rightarrow \infty} A = \frac{1}{2}$ .

We could also use the normalized area  $2A-1$  in which case these limits are 0.

Let  $b=1$ . Then we have a chain. Now it is easy to verify that

$$\alpha_i^+ = \alpha_i^- = 1, i = 1, \dots, d-1$$

$$\alpha_0^+ = 1, \alpha_0^- = 0$$

$$\alpha_d^+ = 0, \alpha_d^- = -1$$

Hence

$$\alpha_i = 0, i = 1, \dots, d-1$$

$$\alpha_0 = 1$$

$$\alpha_d = -1$$

Hence  $X = (\alpha_1, \dots, \alpha_N)$  (normalized since  $\sum_i \alpha_i = 1$ ) equals

$$(1, \underbrace{0, \dots, 0}_{d-1}, -1).$$

Note that  $y=1$ . Further  $A = \frac{d}{d+1}$  as is readily seen. Now  $\lim_{d \rightarrow \infty} A=1$ , contrary to the case  $b$  not 1.

In Egghe (2001), GHT for a chain has been calculated, i.e. GHT for a rooted tree with  $b=1$ . We leave it as an exercise to the reader to calculate the formulae of GHT for a general rooted tree.

## References

- R.A. Botafogo, E. Rivlin and B. Shneiderman (1992). Structural analysis of hypertexts : identifying hierarchies and useful metrics. *ACM Transactions on Information Systems*, 10(2), 142-180.
- W.-K. Chen (1971). *Applied Graph Theory*, Amsterdam : North-Holland.
- P. De Bra (2000). Using hypertext metrics to measure research output levels. *Scientometrics*, 47(2), 227-236.
- L. Egghe (2001). Development of hierarchy theory for digraphs using concentration theory based on a new type of Lorenz curve. *Mathematical and Computer Modelling* (to appear).
- L. Egghe and R. Rousseau (2001). Co-citation, bibliographic coupling and a characterization of lattice citation networks. *Scientometrics* 55(3), 349-361, 2002.
- R.J. Wilson (1972). *Introduction to Graph Theory*. London : Longman.