Ego Citation Networks Considered as Domination Networks

Yuxian Liu¹,², Ronald Rousseau³,*

¹Tongji University Library, Tongji University, Siping Street, Yangpu, Shanghai, CHINA.
²School of History and Archives, Yunnan University, Kunming, CHINA.
³University of Antwerp, Faculty of Social Sciences, Middelheimlaan, Antwerpen, and KU Leuven ECOOM, Naamsestraat, Leuven, BELGIUM.

ABSTRACT

In this article we continue our study of power structures. A dynamic quantitative theory, including the measurement of power or dominance structures is applied to citation networks. A curve somewhat similar to the Lorenz curve for inequality measurement is used. We calculate the D-measure, a normalized measure expressing the degree of dominance in a network. The D-measure of a citation network consisting of the ego (the original article) with n references and m received citations is obtained. When n=m the network is symmetric and the D-measure is 0.5; when the article did not yet receive a citation the D-measure is n/(n+1). When the article has no references then the D-measure is 1/(m+1). A real-world case is described and the evolution of its D-measure over time is shown. Our work is one way of describing an evolving ego-citation network in a quantitative way.

Keywords: Power structure, Network dynamics, Citation graphs, Time evolution, Network properties.

INTRODUCTION

It goes without saying that network theory is an essential part of contemporary science. In the field of Library and information science (LIS) article citation networks, author citation networks, author collaboration networks, bibliographic coupling and co-citation networks are among the best known.[¹] When collecting data and studying networks one may focus on one particular actor and their properties within the corresponding ego network (a network build from the perspective of one actor, the ego) or one may focus on the network as a whole, including all relations of all actors belonging to this network.

Of course, one may also study global properties of ego networks. A typical example of an ego network is White’s description of Eugene Garfield’s research network.[²] In this contribution we study a global property, called dominance, of networks, focussing on citation networks. The included real-world example is an ego network.

This article applies the notion of dominance as introduced in an earlier article.[³] In this earlier work we discussed global dominance as a special network property based on the idea of zero-sum arrays. One may say that the ‘dominance’ idea is a way to operationalize ubiquitous terms such as ‘top’, ‘leading’ or ‘superior’. Practically, we try to contribute to a quantitative description of the changing – global – structure of an evolving citation network. Our approach starts from graph theory and can be applied to real data of any kind. At this level there is no necessity for data to conform to certain modeling aspects such as scaling, self-organization or power law behavior. Experience with such data and resulting numerical values may in the future lead to a way to distinguish between different types of evolutions of citation networks. Parts of this contribution were presented during the ISSI conference in Wuhan (China).[⁴]

Zero-sum arrays and D-curves

In this section we briefly recall the definitions and main results of[³] as these will be necessary to understand the developments presented further on.

Definition: arrays

If X is a (finite) array, i.e. an N-tuple (N > 1), then the j-th element of X is denoted as (X)j = xj, where xj is a real number. In this article components of any array are assumed to be ranked in decreasing order.
Definition: zero-sum array

If \( X = (x_1, \ldots, x_N) \) is a real-valued array, i.e. an \( N \)-tuple, such that \( \sum_{i=1}^{N} x_i = 0 \), then \( X \) is called a zero-sum array. The set of all zero-sum arrays is denoted as \( \mathbb{Z} \).

Constructing a pseudo-Lorenz curve for zero-sum arrays

Definitions and notations

Given a zero-sum array \( X \), we set \( I_0 = \{ \{ i \in \{1, \ldots, N\} \text{ such that } x_i > 0 \} \} \), \( I_0 = \{ \{ i \in \{1, \ldots, N\} \text{ such that } x_i = 0 \} \} \) and \( I_- = \{ \{ i \in \{1, \ldots, N\} \text{ such that } x_i < 0 \} \} \).

As in \(^3\) we assume that \( X \) is not the trivial zero-array, hence \( I_0 \neq \{1, \ldots, N\} \). This implies that \( I_0(X) \) and \( I_0(X) \) are always non-empty, but they may have different numbers of elements. When it is clear about which array we are talking or when it does not matter we simply write \( I_0 \), \( I_0 \) or \( I_- \).

We note that \( \sum_{i \in I_0} x_i = -\sum_{i \in I_-} x_i \). Next we put and \( \Sigma_+ = \sum_{i \in I_0} x_i \).

For each zero-sum array \( X \), we associate a corresponding \( A \)-array, denoted \( A_X \), and equal to \( A_X = (a_1, \ldots, a_N) \). Also \( A_X \) is a zero-sum array. Related to the array \( X \), we will further need the array \( Q_X \), with \( (Q_X)_i = q_i = \sum_{k=1}^{i-1} |a_k| \) where \( |.| \) denotes the absolute value of a number. Clearly \( q_N = 2 \).

Construction of a D-curve

D-curves (D for dominance) of a zero-sum array \( X \), have been introduced in.\(^4\) A D-curve of a zero-sum array \( X \) is defined as the polygonal line connecting the points

\[
(0,0) \rightarrow \left( \frac{1}{N}, a_1 \right) \rightarrow \ldots \rightarrow \left( \frac{i}{N}, \sum_{j=1}^{i} a_j \right) \rightarrow \ldots \rightarrow \left( \frac{|I_+|}{N}, 1 \right) \rightarrow \left( \frac{N-|I_-|}{N}, 1 \right) \rightarrow \ldots \rightarrow \left( \frac{k}{N}, \sum_{j=1}^{k} |a_j| \right) \rightarrow \ldots \rightarrow (1,2)
\]

where \( I_+, k \in I_- \).

This curve can be described as a function, denoted as \( D_X(t) \): for \( t \in [0,1] \), we have:

\[
D_X(t) = \begin{cases} 
\sum_{i=1}^{tN} a_i + N a_{N+1} \left( t - \frac{i}{N} \right), & t \in [\frac{i}{N}, \frac{i+1}{N}), \ i = 1, \ldots, N - \#(I_+)-1 \\
\sum_{i=1}^{tN} |a_i| + N |a_{N+1}| \left( t - \frac{i}{N} \right), & t \in [\frac{i}{N}, \frac{i+1}{N}), \ i = N - \#(I_-), \ldots, N - 1
\end{cases}
\]

where \( \# \) denotes the number of elements in a set. We see that a D-curve is partly concave (namely when the \( a \)'s are positive) and partly convex (the part where the \( a \)'s are negative), as illustrated in Figure 1.

If \( \#(I_+) = N - \#(I_-) \) then the D-curve has no horizontal part. If \( I_0 \neq \emptyset \) then it has a horizontal middle part, at height 1.

An example.

The D-curve of \((6,2,0,0,-1,-1,-2,-4), N=8\), has \( A \)-array (array of \( a \)-values) equal to \( A = \begin{pmatrix} 6 & 2 & 0 & 0 & -1 & -1 & -2 & -4 \end{pmatrix} \), hence connects points with ordinates \( \begin{pmatrix} 0, \frac{6}{8}, \frac{8}{8}, \frac{8}{8}, \frac{8}{8}, \frac{8}{8}, \frac{8}{8}, \frac{8}{8}, \frac{8}{8} \end{pmatrix} = 2 \).

This D-curve is shown in Figure 1.

Definition. Equivalent zero-sum arrays

Zero-sum arrays that lead to the same D-curve are said to be equivalent. The arrays \((5,1,0,0,-2,-4), (10,2,0,0,-4,-8)\) and \((\frac{5}{6}, \frac{1}{6}, 0, 0, -\frac{2}{6}, -\frac{4}{6})\) are equivalent. Equivalent zero-sum arrays of length \( N \) all have the same \( A \)-array. Also \((4,1,0,-5)\) and \((4,4,1,1,0,0,-5,-5)\) are equivalent zero-sum arrays, but with different length.

Figure 1: D-curve of the array \((6,2,0,0,-1,-1,-2,-4)\).
**Partial orders for zero-sum arrays**

Definition: the dominance relation $\leq_d$ in $\mathbb{Z}$

Let $X$ and $Y$ be zero-sum arrays, not necessarily of the same length, then we say that $X$ is $D$-smaller than $Y$ (or $Y$ is $D$-larger than $X$), denoted as $X \leq_d Y$ (or $Y \geq_d X$) if, for each $t \in [0,1]$, $D_x(t) \leq D_y(t)$. $X$ is strictly $D$-smaller than $Y$, denoted as $X <_d Y$, if, for each $t \in [0,1]$, $D_x(t) < D_y(t)$ and there is at least one point $t_i$ hence infinitely many, where $D_x(t_i) < D_y(t_i)$. When $X \leq_d Y$ it is clear that the $D$-curve of $X$ lies nowhere strictly above the $D$-curve of $Y$. The relation $\leq_d$ determines a partial order in the set of all equivalence classes of zero-sum arrays. Formally we write:

$$X \leq_d Y \text{ if and only if } \forall t \in [0,1]: D_x(t) \leq D_y(t)$$

Moreover, $X = Y$ (as equivalence classes) if and only if $\forall t \in [0,1]: D_x(t) = D_y(t)$.

As the dominance relation $\leq_d$ is only a partial order, some arrays cannot be compared: they are said to be intrinsically incomparable.

It is clear that the idea of a $D$-curve is inspired by the idea of a Lorenz curve, but has different properties.\[5-7\]

**Maximum and minimum $D$-curves**

**Maximum $D$-curves**

For fixed $N$ the maximum $D$-curve occurs when $(0,0)$ is linearly connected to the point with coordinates $(1/N,1)$ and then linearly connected to the endpoint $(1,2)$. This $D$-curve corresponds to all zero-sum arrays of the form $X = (s,-t,...,-t)$, with $s, t > 0$ and $s = (N-1)t$.

Considering $N$ as a variable, the line $y = x + 1$, passing through the points $(0,1)$ and $(1,2)$ is an upper bound for all these $D$-maximum $N$-curves.

**Minimum $D$-curves**

For fixed $N$, a minimum $D$-curve is obtained by connecting the origin $(0,0)$ linearly to the point with coordinates $\left( \frac{N-1}{N}, 1 \right)$ and then further linearly to the point $(1,2)$. This minimum $D$-curve corresponds to all arrays of the form $Y = (u,...,u,-v)$, with $u, v > 0$ and $v = (N-1)u$. The line $y = x$, passing through the origin and the point $(1,1)$ is a lower bound for all these minimum $D$-curves. We note that if $N$ increases then the maximum $D$-curve becomes larger too (in the partial order of $D$-curves). Similarly, minimum $D$-curves become smaller.

A measure respecting the dominance relation $\leq_d$ in $\mathbb{Z}$

The area between the $D$-curve and the line $y = x$ respects the $D$ partial order. This area is denoted as $AR_d(X)$. For any zero-sum array this area takes values on the interval $[0,1]$. We will refer to this area as the $D$-measure. The $D$-measure is denoted $AR_d$ and is calculated as:

$$AR_d(X) = \frac{1}{N}\left( \sum_{i=1}^{N} q_i \right) - \frac{N+2}{2N}.$$  

where the $q$-values are the components of the array $Q_x$ with $(Q_{x}) = q_{ij} = \sum_{k=1}^{N} |a_k|$

An example: for $X = (5,2,0,-3,-4)$ the $D$-measure is equal to:

$$AR_d(X) = \frac{1}{5}\left( \frac{5}{7} + \frac{7}{7} + \frac{7}{7} + \frac{10}{7} + \frac{14}{7} \right) - \frac{7}{10} = \frac{37}{70}$$

Maximum $D$-curves have $D$-measures equal to $\frac{N-1}{N}$, while minimum $D$-curves have $D$-measures equal to $\frac{1}{N}$. Although some $D$-curves are intrinsically incomparable, their $D$-values, being positive real numbers are always comparable.

**Applications to directed networks, such as direct citation networks**

**Graph theory**

We briefly recall the basic graph theoretical terminology needed in the sequel and note that in this article the terms graph and network are considered to be synonyms. A directed graph (in short: digraph), denoted $G(V,E)$ consists of a set of vertices or nodes, denoted as $V$ or $V(G)$ and a set of edges or links, denoted as $E$ or $E(G)$. Nodes will be denoted by lower case letters such as $i$ and $j$. An edge is an ordered pair of the form $(i,j)$ where $i$ and $j$ are nodes, hence belong to $V$. Node $i$ is called the initial node and node $j$ is the terminal node of edge $(i,j)$. A directed path, or chain, from node $i$ to node $k$ is a sequences of edges $(v_{p})_{p=1,...,M}$ such that the terminal node of edge $v_{p}$ coincides with the initial node of edge $v_{p+1}$ and such that node $i$ is the initial node of edge $v_{1}$ and node $k$ is the terminal node of edge $v_{M}$. If node $i$ coincides with node $k$ the directed path is a directed circuit or loop. A directed graph is called acyclic or loopless if it does not contain directed circuits. A directed graph is weakly connected if a path exists between any two nodes in the underlying undirected graph. We will always assume that the graphs we study have a finite number of nodes, at least two and are acyclic and weakly connected.

In\[3\] we introduced a local and a global dominance theory. In this contribution, we restrict ourselves to the global theory, in short GDT. In this theory we use arrays of the form $\Sigma = (\sigma_1, \sigma_2, ..., \sigma_N)$, defined as follows:
powerless subordinates. In applications of D-curves to institutes, research groups or scientists as nodes we want to gauge the power structure that is present. The more inequality among nodes with a positive flow the more powerful the order relation is. But also the more even (in the sense of evenness) the nodes with a negative flow, the more powerful the order.

Dynamic aspects of networks and properties of D-curves

When a new measure is proposed one usually derives theoretical properties and explains the possible benefits of using such a measure. This has been done in. Studying dynamic aspects of networks and their corresponding dominance measures is a next and essential step for potential applications in fields such as business management, politics and social interactions. Here we restrict ourselves to applications in citation analysis.

We already know that adding a node in a digraph which dominates the network source, makes this new node the network source, hence it becomes a global dominance node. Linear structures are clear hierarchies but they are not interesting in the context of power structures: such structures are always intrinsically incomparable and because of their symmetry their D-measures are always equal to 0.5. In this article we study some aspects of evolving ego-citation networks.
An ego citation network

A simple example

An author chooses which articles to include in his/her reference list. For this reason one can consider a citing article to be in a dominating position with respect to the article which receives a citation. We note though that scientifically one may argue that the cited article is the superior one as the citing article recognizes it as an authority. From the abstract point of view of the network research presented here it does not really matter which point of view is taken (it is just a matter of reversing the arrows in a directed network), but, of course, socially and emotionally it does.

We consider an article that has four references and which receives step by step more and more (direct) citations, making the structure more and more top-heavy. In this simple example we only consider the position of the original article with respect to its references and articles citing it, neglecting other nodes in the citation network. The different steps are shown in Table 1. Such a citation network always begins with a maximum D-curve.

A general formula

It is possible to derive a general formula for the global D-measure in the case that the original article has n references and received m citations, hence N = n+m+1. Note that possible relations between cited and cited, citing and citing or cited and citing articles are not taken into account in this calculation. This case is illustrated in Figure 4.

The corresponding array is:

\[
X = \left\{ \frac{1+2n}{m \times n} \times \ldots \times \frac{1+2n}{m \times n}, -\frac{m-n}{m(1+2n)} \times \ldots \times -\frac{m-n}{m(1+2n)} \right\}
\]

If m \geq n then the |A|- array (array of absolute values of the A-array) is:

\[
\frac{1+2n}{m \times n} \times \ldots \times \frac{1+2n}{m \times n}, -\frac{m-n}{m(1+2n)} \times \ldots \times -\frac{m-n}{m(1+2n)}
\]

and the Q-array becomes:

\[
\begin{align*}
AR_D(X) &= \left( \frac{1}{N} \sum_{i=1}^{N} q_i \right) - \frac{N}{2} \\
&= \frac{1}{m+n+1} \left( \frac{m+1}{2} + \sum_{j=1}^{m} \left( \frac{1}{m(1+2n)} \times \ldots \times \frac{1}{m(1+2n)} \right) \right) \\
&\quad + \sum_{k=1}^{n} \left( \frac{1}{m(1+2n)} \times \ldots \times \frac{1}{m(1+2n)} \right)
\end{align*}
\]

Table 1: A short citation history of a fictitious article with four references.

<table>
<thead>
<tr>
<th>Number of citing articles</th>
<th>N</th>
<th>Array</th>
<th>Value of the global D-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>(4,-1,-1,-1,-1)</td>
<td>0.8</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>(9,3,-3,-3,-3)</td>
<td>0.708</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>(9,9,2,-5,-5,-5)</td>
<td>0.621</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>(9,9,9,9,0,-9,-9,-9)</td>
<td>0.5</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>(9,9,9,9,9,9,4,-17,-17,-17)</td>
<td>0.357</td>
</tr>
</tbody>
</table>

Figure 4: The general case: an article with n references and m received citations.

One can easily check that, if m = n, the value of AR_D(X) is equal to 0.5. If n = 0, corresponding with an article without references, then the result is 1/(m+1). This case clearly corresponds with a decreasing power structure for increasing m.
If now \( m \leq n \) then the positive part is equal to \( n (2m+1) \) and

\[
AR_c(X) = \frac{1}{N} \left( \sum_{k=0}^{N} n_k \right) - \frac{N+2}{2N}
\]

\[
= \frac{1}{m+n+1} \left( \sum_{j=0}^{n} \frac{j(1+2m)}{n(1+2m)} + \sum_{k=2}^{n} \frac{1}{n} \right) - \frac{m+n+3}{2(m+n+1)}
\]

\[
= \frac{1}{m+n+1} \left( \frac{m(2n+1) + m(6n^2 + 8n + 1) + 3n(n+1)}{2(n(1+2m))} \right) - \frac{m+n+3}{2(m+n+1)}
\]

\[
= \frac{m^2 + m(4n^2 + n + 1) + 2n^2}{2n(n+m+1)(2m+1)}
\]

Also, here the expression becomes equal to 0.5 when \( m = n \). If \( m = 0 \), corresponding with an article that has not been cited yet, then the result is \( n/n(n+1) \).

**A real-world example**

As a real-world example we consider the article[9] as the ego in a citation network. This article’s full bibliographic data are:

<table>
<thead>
<tr>
<th>Cumulative number of citing articles - year</th>
<th>N</th>
<th>X-array</th>
<th>Value of the D-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 - 2009</td>
<td>18</td>
<td>( 17, -1, \ldots, -1 )</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 17 \text{times} )</td>
<td></td>
</tr>
<tr>
<td>1 - 2010</td>
<td>19</td>
<td>( 39, 16, -3, -4 )</td>
<td>0.881</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 13 \text{times} 4 \text{times} )</td>
<td></td>
</tr>
<tr>
<td>4 – 2011, 2012, 2013</td>
<td>22</td>
<td>( 39, 37, 36, 13, -9, -10, -11, -12 )</td>
<td>0.772</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 11 \text{times} 4 \text{times} )</td>
<td></td>
</tr>
<tr>
<td>5 - 2014</td>
<td>23</td>
<td>( 93, 39, 37, 36, 10, -14, -15, -17, -20 )</td>
<td>0.761</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 11 \text{times} 4 \text{times} )</td>
<td></td>
</tr>
<tr>
<td>6 – 2015, 2016</td>
<td>24</td>
<td>( 93, 39, 37, 36, 9, -16, -17, -19, -23 )</td>
<td>0.732</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 11 \text{times} 4 \text{times} )</td>
<td></td>
</tr>
<tr>
<td>7 - 2017</td>
<td>25</td>
<td>( 93, 39, 37, 36, 36, 8, -18, -19, -20, -22, -26 )</td>
<td>0.704</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 10 \text{times} 4 \text{times} )</td>
<td></td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENT

The authors thank anonymous reviewers for thoughtful observations. They further thank Leo Egghe (University Hasselt, Belgium) for helpful discussions. This work was supported by NSSFC via 71173154, National Social Sciences Foundation of China. Finally, we also thank Mats Benner, Joshua Eykens and Raf Guns for helpful information.

CONFLICT OF INTEREST

The authors declare that their is no conflict of interest.

ABBREVIATIONS

GDT: Global Domination Theory.

REFERENCES


Appendix: Articles in the real-world citation network example EGO


18. The Stimulate-6 Group. The Hirsch index applied to topics of interest to developing countries. First Monday. 2007;12(2). http://www.firstmonday.org/issues/issue12_2/stimulate/Articles citing the EGO (7 articles, mentioned chronologically).