RESEARCH

On some properties of generalized Lexis diagrams

Tim Riffe* and Jonas Schöley

*Correspondence: riffe@mpidr.mpg.de
Full list of author information is available at the end of the article
†Equal contributor

Abstract

Time identities are relationships implied by events and the durations between them, such at the well-known age-period-cohort relationship. We describe the construction and structure of such identities.

Keywords: age structure; formal demography; data visualization; age period cohort

1 Introduction

Population processes are temporally structured. Sometimes we structure population stocks or flows by more than one element of time at once. Elements of time may include the moment of observation, the time points of events such as birth or death, or the time passed between events. Time measures can be categorized into two basic types: events and durations. Events include any measure at a point *point* in time. This notion includes period itself, which one may imagine for our purposes as an infinite set of points. This is not to be confused with the common notion of period as a bounded duration of observation. Durations are time differences between pairs of events: Chronological age is an example of a duration, A = P − C, for example, and many other durations may result from taking the time differences between arbitrarily paired events. For example, adding the event of marriage to the mix, M, results in several further implied time measures: age at marriage = M − C, time until marriage (time married) = M − P, if P ¡ M (M ¿ P), and so forth.

2 Constructing temporal identities.

We begin this exposition by describing the construction of temporal identities, largely reproduced in the present section from Riffe et al. (2017). In the following we describe event-based time frameworks in terms of vector spaces which, via linear transformation, relate the timing of events with durations between events.

Definition 2.1 Let $\boldsymbol{p} = (p_1, \ldots, p_n)^\top \in \mathbf{R}^n$ be a vector of n events or points in time with $n \geq 2$. A corresponding vector of durations $\boldsymbol{d} \in \mathbf{R}^m$ is composed by elements of the form $d_{i,j} = p_j - p_i$ for $i = 1, \ldots, n-1$, $j = 2, \ldots, n$ and $j > i$.

The vector of events $\boldsymbol{p}$ can be ordered in an arbitrary way as long as the same elements in $\boldsymbol{p}$ correspond to the same type of event for all observations. A consequence of this is that durations may be either negative or positive depending on the ordering of events over the life course.

Proposition 2.1 Given a vector of events $\boldsymbol{p} = (p_1, \ldots, p_n)^\top \in \mathbf{R}^n$, the dimension of the corresponding vector of durations $\boldsymbol{d} \in \mathbf{R}^m$ is $m = n(n-1)/2$.

Proof By definition, each element of $\boldsymbol{d}$ is formed by two different elements of $\boldsymbol{p}$. Therefore, the length of $\boldsymbol{d}$ is the number of combinations of 2 different elements from a set of size n, such that the order of selection does not matter. From combinatorial theory, it is well known that this value is given by the binomial coefficient $\binom{n}{2} = \frac{n!}{2!(n-2)!} = n(n-1)/2$. □

Proposition 2.2 For any vector of events $\boldsymbol{p} = (p_1, \ldots, p_n)^\top \in \mathbf{R}^n$, there is always a linear transformation $f : \mathbf{R}^n \to \mathbf{R}^m$ that provides a corresponding vector of durations $\boldsymbol{d} \in \mathbf{R}^m$.

Proof The existence of f is a direct consequence of Definition 2.1, given that all the elements of $\boldsymbol{d}$ are a linear combination of elements of $\boldsymbol{p}$. □

Corollary 2.2.1 Given $\boldsymbol{p} = (p_1, \ldots, p_n)^\top \in \mathbf{R}^n$, suppose that $\boldsymbol{d} = (p_2 - p_1, \ldots, p_n - p_1, p_3 - p_2, \ldots, p_n - p_2, \ldots, p_n - p_{n-1})$. Then, the linear transformation $f : \mathbf{R}^n \to \mathbf{R}^m$ that yields $\boldsymbol{d}$ from $\boldsymbol{p}$ is defined by the $m \times n$ matrix

$$\boldsymbol{X}_{(m\times n)} = \left(\begin{array}{c|c|c|c|c}
-1 & & & & \\
\vdots & & I_{n-1} & & \\
-1 & & & & \\ \hline
0 & -1 & & & \\
\vdots & \vdots & & I_{n-2} & \\
0 & -1 & & & \\ \hline
 & & \cdots & & \\ \hline
0 & \cdots & 0 & -1 & I_2 \\
0 & \cdots & 0 & -1 & \\ \hline
0 & \cdots & 0 & -1 & 1
\end{array}\right), \tag{1}$$

such that $\boldsymbol{d} = \boldsymbol{X} \times \boldsymbol{p}$, and where I_k denotes the $k \times k$ identity matrix.

These results imply that given an arbitrary set of $n \geq 2$ points in time, it is always possible to calculate the durations between any pair of these points. However, note that matrix $\boldsymbol{X}$ in (1) yields a vector of durations $\boldsymbol{d} \in \mathbf{R}^m$ whose elements are sorted in an arbitrary way. The following statement may be relevant in this regard.

Proposition 2.3 Given a vector of events $\boldsymbol{p} = (p_1, \ldots, p_n)^\top \in \mathbf{R}^n$, the corresponding vector of durations $\boldsymbol{d} \in \mathbf{R}^m$ is unique, irrespective of the sorting of its elements.

Proof Let's suppose that $\boldsymbol{d}^{\mathbf{1}}$ and $\boldsymbol{d}^{\mathbf{2}}$ are two different vectors of durations corresponding to the same vector of events $\boldsymbol{p} \in \mathbf{R}^n$. Provided that $\boldsymbol{d}^{\mathbf{1}}$ and $\boldsymbol{d}^{\mathbf{2}}$ are finite and, by definition, both have dimension m and are formed by the same combinations of elements of $\boldsymbol{p}$, it will always be possible to re-arrange the elements of $\boldsymbol{d}^{\mathbf{2}}$ in the same order as $\boldsymbol{d}^{\mathbf{1}}$ such that $\boldsymbol{d}^{\mathbf{1}} = \boldsymbol{d}^{\mathbf{2}}$. □

This last proposition allows considering $\boldsymbol{X}$ as the matrix defining the linear transformation between points and durations. Given a vector $\boldsymbol{p}$ and the corresponding $\boldsymbol{d} = \boldsymbol{X} \times \boldsymbol{p}$, any differently sorted vector of durations would be obtained by swapping the rows of $\boldsymbol{X}$. Further, note that $\boldsymbol{X}$ does not have an inverse matrix, and therefore there is no linear transformation from durations to events. This is intuitively straightforward if one thinks that two vectors of events can yield the same vector of durations. In other words, a particular vector of durations can come from infinite different vectors of points in time. For instance, using $\boldsymbol{X}$, the vectors of events $\boldsymbol{p}^{\mathbf{1}} = (1, 2, 3)$ and $\boldsymbol{p}^{\mathbf{2}} = (2, 3, 4)$ both yield $\boldsymbol{d} = (1, 2, 1)$.

The relationship between events and durations can be systematically represented in a series of timelines and graphs that may better guide intuition. The joint relationship between events and durations is more explicit and more compact in a graph representation. As introduced in the following definition, the total number of time measures implied by a set of n events and the corresponding durations is $n + m = n + n(n-1)/2 = n(n+1)/2$.

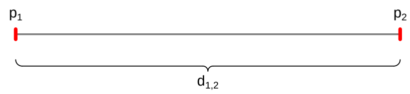
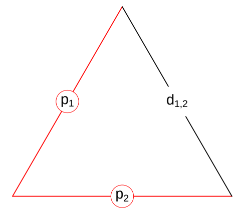
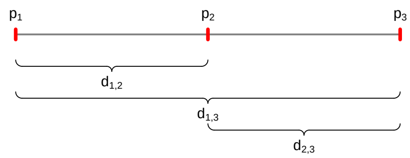
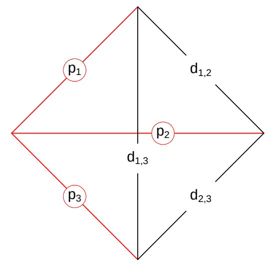
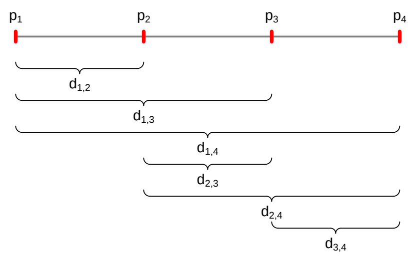
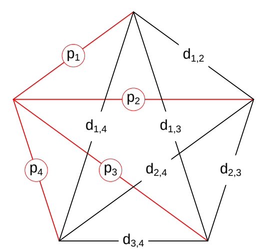
Definition 2.2 Given a vector of events $\boldsymbol{p} = (p_1, \ldots, p_n)^\top \in \mathbf{R}^n$, $n \geq 2$, and the corresponding vector of durations $\boldsymbol{d} \in \mathbf{R}^m$, we define the graph of time measures $\boldsymbol{G}$ as the graph with $n + m = n + n(n+1)/2$ edges labelled by $(\boldsymbol{p}, \boldsymbol{d}) \in \mathbf{R}^{n(n+1)/2}$ such that the relationships in Definition 2.1 are preserved.

Table 3 displays a timeline and a graph for two, three, and four event sets. The central column shows timelines, a familiar linear representation of time, with events marked with red ticks labelled with $p_1 \ldots p_n$. Durations span each of the m possible event dyads and are drawn below the main timeline as curly braces labelled with $d_{1,2} \ldots d_{n-1,n}$. The right column of Table 3 draws the corresponding graph with a total of $n+1$ vertices and $n+m = n(n+1)/2$ edges for the elements of both $\boldsymbol{p}$ and $\boldsymbol{d}$. All events of $\boldsymbol{p}$ connect to a single vertex, and event edges are indicated in red with red-circled labels. In this rendering, each triangle formed by three mutually connecting edges represents a triad identity. The top row $n = 2$ consists in a single identity. Three and four events imply a total of four and ten triad identities, respectively, and in general a given higher order identity will yield $\binom{n+1}{3}$ triad identities. We call this a temporal plane graph because the triangle resulting from any given triad sub-identity can be extended over all valid values of its time measures to form a temporal plane, as of the diagrams in a previous section (not in excerpt). The dimensionality of the extended diagram of a given identity follows from the number of events from which the identity is derived: $n = 2$ produces a two-dimensional diagram, $n = 3$ produces a 3-dimensional diagram, and so forth.

Definition 2.3 We define $P \subseteq \mathbf{R}^n$ as the vector-space (event-space) spanning all possible values vector $\boldsymbol{p}$ may take, and $D \subseteq \mathbf{R}^m$ as the vector space spanning all possible instances of the duration vector $\boldsymbol{d}$.

Just like the APC diagram allows for *all possible* combinations of period, cohort and age we may consider the vector space P spanning all possible instances of $\boldsymbol{p}$. The calculation of durations between events as described in Proposition 2.2 can then be understood as a linear transformation from a vector space P whose bases represent events to a duration vector space D whose bases represent durations.

Table 1: Event-duration timeline and graph for two, three, and four event sequences.

nr. events	timeline	graph
$n = 2$		
$n = 3$		
$n = 4$		

3 Properties of higher order time identities

Here we list and prove some of the properties of higher-order identities, such as the number of ways they can be derived, the conditions for doing so, the number of event and duration measures they contain, and the number, size, and composition of sub-identities.

Proposition 3.1 *(Work in progess)* Something that states that you can transform $P^n \leftrightarrow M^n$ where the basis vectors of M are a mixture of point dimensions and duration dimensions.

Definition 3.1 **g** Let $\mathbf{g} = \{p_{i=1,\dots,n}, d_{k=1,\dots,m}\}$

Proposition 3.2 There are $b = (n+1)^{(n-1)}$ many ways to choose n elements out of $\mathbf{g}$ whoose linear combination yields the remaining m elements of $\mathbf{g}$.

Proof This is a case of Cayley's formula (**?**), a result from graph theory which gives the number of possible trees on $k = n + 1$ vertices, $k^{(k-2)}$. In our case, the fully conected graph $\mathbf{K}$ with edges defined by the elements of $\mathbf{g}$ according to the third column of Table 1, is complete. By Cayley's fomula, the number of minimal spanning trees on $\mathbf{K}$ is equal to $k^{(k-2)}$. Four different proofs of this result are given in **?**. The key is to realize that a minimal spanning tree (MST) on a complete graph will have $k - 1$ edges, connected to each other and all k vertices. As such, the remaining possible edges are linear combinations of any given MST. $\square$

Corollary 3.2.1 Each set of n elements from $\mathbf{g}$, $\mathbf{b}'$ whose linear transformation yields the remaining m elements of $\mathbf{g}$ includes at least one element of $\mathbf{p}$.

Proof One of the vertices of $\mathbf{K}$, say the k^{th} vertex, is connected only to edges labelled by the elements of $\mathbf{p}$. Since a spanning tree of $\mathbf{K}$ must connect to this vertex to fully connect $\mathbf{K}$, all b valid spanning trees must contain at least one edge labelled by an element of $\mathbf{p}$. $\square$

Definition 3.2 **G** Let's define $\mathbf{G}$ as the identity implied by $\mathbf{g}$ whose graph is $\mathbf{K}$.

Proposition 3.3 An identity $\mathbf{G}$ implies a total of $\binom{n+1}{3}$ triad sub-identities.

Proof Any set of three vertices from $\mathbf{K}$ forms a complete subgraph, and any complete subgraph implies an identity between its labelled edges. $\mathbf{K}$ has $n+1$ vertices, and therefore there are $\binom{n+1}{3}$ ways to select three vertices from $\mathbf{K}$, hence $\mathbf{G}$ implies the same number of triad subidentities. $\square$

Corollary 3.3.1 Of the $\binom{n+1}{3}$ triad identies implied by $\mathbf{G}$, $\binom{n}{2}$ contain exactly two events and one duration.

Proof This follows by noting that the n event-labelled edges in $\mathbf{K}$ connect to a single vertex. Selecting any two of these n event-labelled edges implies a tree on three vertices, whose full connection implies a triad identity composed of the two event edges and one duration edge defined as the time-difference of the former two. There are $\binom{n}{2}$ ways to select two of of the n event edges. $\square$

Corollary 3.3.2 For $n \geq 3$, of the $\binom{n+1}{3}$ triad identies implied by $\mathbf{G}$, $\binom{n+1}{3} - \binom{n}{2} = \binom{n}{3}$ are composed of exactly three durations.

Proof This is equivalent to deleting the vertex k^{th} from $\mathbf{K}$, the vertex that connects only to event-labelled edges, which is constructed following the middle column of graphs from Table 1 with vertex labels ignored. This graph has n total vertices, and any set of 3 vertices implies an identity between its three labelled edges, which in this case by definition can only consist of durations. $\square$

For example, in the demographic time identity there are $n = 3$ event measures. Thus of the $\binom{4}{3} = 4$ triad identities implied $\binom{3}{3} = 1$ of these identities consist in durations only (TAL). Notice that the measures T and A change over the lifecourse of an individual, whereas their sum L is fixed.

Definition 3.3 $\mathbf{d}_t$ For $\mathbf{p}$ that include period itself, let $\mathbf{d}_t$ be the set of duration time measures that change over the life course and $\mathbf{d}_f$ consist in those durations that are fixed attributed of an individual. By definition, $\mathbf{d}_t \cup \mathbf{d}_f = \mathbf{d}$.

Corollary 3.3.3 For $n \geq 4$ and For $\mathbf{p}$ that include period itself, of the $d' = \binom{n}{3}$ triad identies whose edges are labelled only by the elements of $\mathbf{d}$, $d^t = \binom{d'}{2}$ of these identities consist in exactly two elements of $\mathbf{d}_t$ and one element of $\mathbf{d}_f$, while $d' - d^t$ of the duration-only triad identities consist in relationships between three elements of $\mathbf{d}_f$.

Proof $n-2$ of the edges in K are labelled with the elements of $\mathbf{d}_t$, and these all connect to the same vertex. There are therefore $\binom{n-2}{2}$ ways to form triad identities with them. The third element of each of these identities cannot connect to the same vertex, and so must be a member of $\mathbf{d}_f$. □

Proposition 3.4 In general, the number of subidentities of size h in $\mathbf{G}$ is equal to $\binom{n+1}{n+1-h} \quad \forall\, h \leq n$.

Proof Vertex deletion on a complete graph results in a complete subgraph. Therefore, the number of possible complete subgraphs with h vertices is a function of the number of ways that $n+1-h$ vertices can be deleted from $\mathbf{K}$, which is $\binom{n+1}{n+1-h}$. The labelled edges of each possible complete subgraph defined in this way represent subidentities. □

For example, from the tetrahedral graph in Fig. **??**, we may delete the vertex that joins the edges labelled A, T, and P, which in effect deletes these edges, leaving us with the CDL identity.

Corollary 3.4.1 Each time measure in $\mathbf{G}$ is a member of $n-1$ triad subidentities.

Proof In the graph $\mathbf{K}$, an edge labelled by a given time measure connects to two of the $n+1$ vertices. A full connection to any other vertex yields a complete subgraph with three edges, representing a triad identity. There are $n-1$ remainign vertices that can be connected to, ergo the given edge is a member of $n-1$ triad subidentities. □

Acknowledgements
We wish to thank Francisco Villavicencio for comments that improved our exposition.

Author details

References

Riffe, T., J. Schöley, and F. Villavicencio (2017). A unified framework of demographic time. *Genus 73*(1), 7.