Математички Билтен
14 (XL)
1990 (95-104)
Скопје, Југославија

GENEALOGICAL TREES AND THEIR GRAPHICAL REPRESENTATION

Katerina Čundeva

Abstract. Family data are summarized and observed in a pedigree chart. Algorithms for their computerized graphical representation are given.

In common usage, the vertices shown as squares correspond to male and those shown as circles to female members of an observed family. An arc exists between two persons with a consanguinity of first degree (parent-child, siblings) or between biological spouses.

All the relatives of one generation are usually represented by parallel strings, as shown in Figure 1 (McKusick, 1969) or by concentric circles, (Figure 2; Vogel, Motulsky, 1979). This article deals with the graphical representation of pedigree charts with parallel generations.

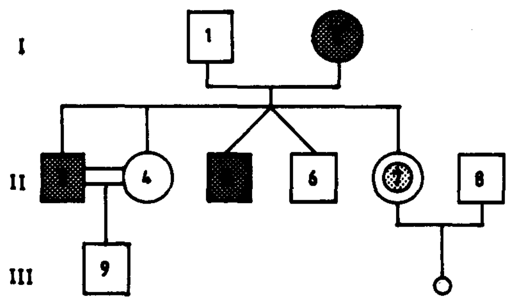


Figure 1. Pedigree chart with parallel generations

The first problem which appears when heredity analysis is performed by a computer is the method of coding and storing genealogical trees. This task can simply be performed by forming consanguinity matrices (Čundeva, 1987). The i-th row and column in such matrices are strings of consanguinity codes belonging to the i-th member M_i. The vector-row contains the codes corresponding to the M_i's relationship towards all the other members of the family, while the vector-column contains all the codes of relationship towards M_i. When a new person M_j is added to the tree, after determining his ordinal number and his generation, the j-th row is produced by using the vector-row of a single, previously coded M_j's relative. There is a 1-1 mapping between the j-th row and the j-th coloumn. When the coding process is finished the chart can be drawn.

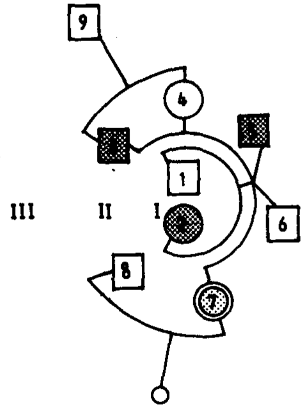


Figure 2. The same pedigree with concentric generations

Vertex position in a pedigree with parallel generations

All coordinates given in the forthcoming algorithms are valid for a coordinate system with zero in the upper left corner and graphical resolution RH*RV. If the coordinate zero is in the lower left corner, then the ordinates Y have to be drawn as RV-Y.

In a pedigree with G generations, the graphical output is divided into horizontal stripes with a width V of

$$V = RV \text{ div } G \qquad (1)$$

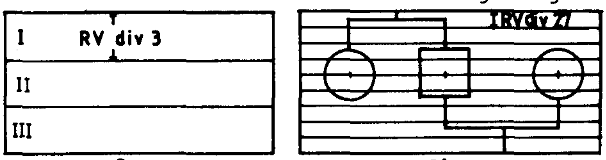
points (Fig. 3a). Each stripe is then subdivided into 9 equal parts with a width $\underline{R}$ given in (2),

$$\underline{R} = \underline{RV} \text{ div } (9*\underline{G}) \qquad (2)$$

In order to obtain the best visibility, the centres of vertices (center of symmetry) are positioned at the fourth line (Fig. 3b). The sides of squares representing males are twice as large as the radius of the female circles called Radius.

$$\underline{Radius} = \min \{(3*\underline{R}) \text{ div } 2,\ \underline{RH} \text{ div } (3*\underline{A})\} \qquad (3)$$

where $\underline{A}$ is the number of members in the largest generation.



3a 3b

Figure 3. Horizontal division in a family with three generations

In the beginning, the vertical distance between the vertices belonging to two neighbouring generations is

$$\underline{Step} = (\underline{RV} - 2*\underline{G}*\underline{Radius}) \text{ div } (2*\underline{G}) \qquad (4)$$

The arcs representing marriage └┘, relationship siblings ┌┐ or parent-child | are $2*\underline{R}$ high.

Now, the horizontal coordinates of the vertices have to be found. A priori, the $\underline{i}$-th person in the $\underline{j}$-th generation that has $\underline{A}_j$ members has its centre in the point with coordinates

$$((2*\underline{i}-1+\underline{A}-\underline{A}_j)*(\underline{Step}-\underline{Radius}),\ 4+9*(\underline{j}-1)*\underline{V} \qquad (5)$$

The distance between two spouses with consecutive ordinal numbers depends on the width of their $\underline{F1}$ (filial) generation. If they have no children, a single unmarried child and at most two unmarried children or a child with one spouse, then the situacion can remain unchanged. But, if they have $\underline{c}$ children, and their children have s spouses, where p + s > 2, then their correspon-ding vertices have to be more separated then previously. This separation implies translations in the parental generation, and furthermore, in the preceding generations. So, a set S of com-mon children and spouses of these children belonging to the same generation has to be constructed. Its cardinal number k(S) can be calculated as:

1. k(S) = 0;
   pl = 0;

2. the first parent P1 has an ordinal number pl;
   -if pl overflows the ordinal numbers in P1's generation, then go to 9;

3. if code A is found at the pl+1-st position in P1's, then the second parent has an ordinal number pl+1;
   -else, pl = pl+1, go to 1;

4. c = 1;

5. if c overflows the family size, then pl = pl+1, go to 1;

6. at the position c in the both vector-rows corresponding to P1 and P2, code C is found, the child C has an ordinal number a, go to 7;
   -else, a = a+1, go to 5;

7. in C's vector-row, among the codes corresponding to the members of C's generation, A is found d times, so k(S) = k(S)+d;

8. a = a+1, go to 5;

9. END.

If S is an empty set, then the situation remains unchanged.

If K(S)=1, the only unmarried child is translated to lie between its parents:

$$T(I_c, J_c) = T(I_f + I_m) \text{div } 2,\ J_c) \qquad (6)$$

where (I_c, J_c) are the child's, (I_f, J_f) its father's and (I_m, J_m) its mother's coordinates. All persons with a higher ordinal number in the child's generation are translated Step+2*Radius points:

$$I > I_c:\ T(I,J) = T(I+Step+2*Radius, J) \quad (7)$$

If $K(S) = 2$ then the child c with a smaller ordinal numer is translated under its parent p with a smaller ordinal number. The others are positioned using the relation (8)

$$I > I_d:\ T(I,J) = T(I,J)+T(I_p,J_p)-T(I_c,J_c) \quad (8)$$

In the most general case, when $K(S) = s$, the first movement occurs in the parental generation. If the father f is before the mother m then their new coordinates are:

$$T(I_m,J_m) = T(I_m+(Radius*(s-2))\ \mathrm{div}\ 2,\ J_m) \quad (9)$$

$$T(I_f,J_f) = T(I_f-(Radius*(s-2))\ \mathrm{div}\ 2,\ J_f) \quad (10)$$

If the parental generation is J-th, then the second movement (10) initializes translation in the previous generations, as given in (11) and (12)

$$I \le I_f:\ T(I,J) = T(I-(s-2)*(Step-4*s)\ \mathrm{div}\ 2,\ J) \quad (11)$$

$$I \le I_m:\ T(I,J) = T(I+(s-2)*(Step-4*s)\ \mathrm{div}\ 2,\ J) \quad (12)$$

If the mother m has a smaller ordinal number, then I_f and I_m reverse their roles and the relations from (9) to (12) are repeated.

The process of positioning the vertices is finished. Depending on the person's sex, squares and circles can be drawn. The left boundary of each vertex is I-Radius, the right I+Radius, the upper J+Radius and the lower J-Radius, where (I,J) are the coordinates of its centre.

Connecting pedigrees with parallel generations

The first lines that can be added to the picture are the lines of marriage between spouses I_1 and I_2 with consecutive ordinal numbers. If none of them has children with other partners the segments T_1T_2, T_2T_3 and T_3T_4 (Fig. 4) are immediately added.

$$\underline{T}_1 = \underline{T}(\underline{I}_1,\ \underline{J}_1+\underline{Radius})$$
$$\underline{T}_2 = \underline{T}(\underline{I}_1,\ \underline{J}_1+\underline{Radius}+2*\underline{R}) \qquad (13)$$
$$\underline{T}_3 = \underline{T}(\underline{I}_2,\ \underline{J}_2+\underline{Radius}+2*\underline{R})$$
$$\underline{T}_4 = \underline{T}(\underline{I}_2,\ \underline{J}_2+\underline{Radius})$$

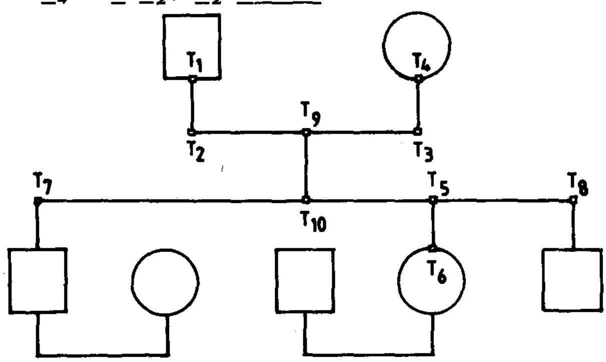


Figure 4. Connection lines in a pedigree with parallel generatior

If one of the partners, e.g. $\underline{I}_1$ has another spouse $\underline{I}_3$, then $\underline{T}_1$ and $\underline{T}_2$ are transformed into:

$$\underline{T}_1 = \underline{T}(\underline{I}_1+3*\underline{a},\ \underline{J}_1+\underline{Radius}-2*\underline{s})$$
$$\underline{T}_2 = \underline{T}(\underline{I}_1+3*\underline{a},\ \underline{J}_1+\underline{Radius}+2*\underline{R}) \qquad (14)$$
$$\underline{a} = \mathrm{sign}(\underline{R}(\underline{I}_2)-\underline{R}(\underline{I}_3))$$
$$\underline{s} = 1 \text{ if } \underline{I}_1 \text{ is a woman}$$

where $\underline{R(I)}$, $\underline{I}\in\{\underline{I}_2,\underline{I}_3\}$ is $\underline{I}$'s ordinal number.

If $\underline{I}_1$ and $\underline{I}_2$ belong to same generation, but their ordinal numbers are not consecutive, then $\underline{T}_2$ and $\underline{T}_3$ are increased by 3$\underline{b}$.

$$\underline{T}_2 = \underline{T}(\underline{I}_2,\ \underline{J}_2+\underline{Radius}+2*\underline{R}+3*\underline{b})$$
$$\underline{T}_3 = \underline{T}(\underline{I}_3,\ \underline{J}_3+\underline{Radius}+2*\underline{R}+3*\underline{b}) \qquad (15)$$
$$\underline{b} = (\underline{R}(\underline{I}_2)-\underline{R}(\underline{I}_3)) \text{ div } 7$$

With this correction their line of marriage passes under al the others. When the spouses have very distant ordinal numbers, then $\underline{b}$ avoids the potential intersections.

When two spouses belong to different generations, then the coordinates $\underline{T}_1$, $\underline{T}_2$, $\underline{T}_3$ and $\underline{T}_4$ are changed into:

$$\underline{T}_1 = \underline{T}(\underline{I}_1+3*\underline{a},\ \underline{J}_1+\underline{Radius}-4)$$
$$\underline{T}_2 = \underline{T}(\underline{I}_1+3*\underline{a},\ \underline{J}_1+\underline{Radius}+2*\underline{R})$$
$$\underline{T}_3 = \underline{T}(\underline{I}_1,\ \underline{J}_1+\underline{Radius}+2*\underline{R}+3*\underline{b}) \quad (16)$$
$$\underline{T}_4 = \underline{T}(\underline{I}_2,\ \underline{J}_2+\underline{Radius}+2*\underline{R}+3*\underline{b})$$
$$\underline{a} = \mathrm{sign}(\underline{R}(\underline{I}_1)-\underline{R}(\underline{I}_2))$$

where $\underline{I}_1$ is the spouse from the older generation.

The children are joined according to the following algorithm:

1. all the members who are not in the first generation get the marker *;
2. w is the lowest ordinal number in the second generation;
3. if w overflows the number of members, go to 10;
4. if w has a marker +, then the child has already been connected, so w=w+1, go to 3;
5. Henseforth, W is the w-th person in the family. All W's siblings, together with W make the set $\underline{S}_1$. Their marker * is changed into +. The highest ordinal number in $\underline{S}_1$ is q, $\underline{k}=\underline{K}(\underline{S}_1)$;
6. The lines towards the parental generation are drawn:

$$\underline{T}_5 = \underline{T}(\underline{I},\ \underline{J}-\underline{Radius}-2*\underline{R}-3*\underline{c})$$
$$\underline{T}_6 = \underline{T}(\underline{I},\ \underline{J}-\underline{Radius}) \quad (17)$$
$$\underline{c} = (\underline{q}-\underline{r})\ \mathrm{div}\ (2*\underline{k}+1)$$

7. The siblings from $\underline{S}_1$ are connected with segments $\underline{T}_7\underline{T}_8$:

$$\underline{T}_7 = \underline{T}(\underline{I}_r,\ \underline{J}_r-\underline{Radius}-2*\underline{R}-3*\underline{c})$$
$$\underline{T}_8 = \underline{T}(\underline{I}_q,\ \underline{J}_q-\underline{Radius}-2*\underline{R}-3*\underline{c}) \quad (18)$$

where $(\underline{I}_r,\underline{J}_r)$ and $(\underline{I}_q,\underline{J}_q)$ are the coordinates corresponding to the first and the last member in $\underline{S}_1$.

8. Henceforth, the parents' coordinates are $(\underline{I}_1,\underline{J}_1)$ and $(\underline{I}_2,\underline{J}_2)$. The parental and fillial generation are joined with $\underline{T}_9\underline{T}_{10}$:

$$\underline{T}_9 = ((\underline{I}_1+\underline{I}_2) \text{ div } 2,\ \underline{J}_1+\underline{Radius}+2*\underline{R}+3*\underline{b}) \qquad (19)$$
$$\underline{T}_{10} = ((\underline{I}_1+\underline{I}_2) \text{ div } 2,\ \underline{J}_1-\underline{Radius}+2*\underline{R}+3*\underline{b})$$

9. $\underline{w} = \underline{w}+1$, go to 3,

10. END.

The given algorithm can be easily adapted for pedigree charts with concentric generations. The horizontal and vertical coordinates are transformed into angles and arcs. So, the new horizontal resolution HV becomes 360, while the vertical RV will be the half of the lower resolution.

All the relations are valid. A point with coordinates (I,J) is represented as (RV*cos(I/360),J*sin(I/360)). The connections $\underline{T}_i\underline{T}_j$ ($\underline{i} \le 10$, $\underline{j} \le 10$) are drawn as arcs with a centre in the middle of the graphical output.

Conclusion

The algorithms for software analysis depend on the type of genetic research. If a single genetic characteristic is observed, then the unaffected persons are given with bright, the affected with dark and the carriers with shadowed symbols (Fig. 1, Fig. 2).

The associations and syndromes are shown with divided symbols. Each part of the symbol stands for one attribute. Usually each anomaly has its own colour, so the problem is how to obtain a clear monochrome pedigree chart.

Usually there is no need to show the complete pedigree chart. Sometimes heredity is observed only through direct descendents, without paying attention to their partners. Another reduction results from the fact that some characteristics appear only in particular branches of the tree. This means that the pedigree chart has to be reorganized. The main forthcoming task is the way of showing subtrees without performing new consangunity matrices.

These algorithms are developed to give a facility of getting, adding and observing data for genealogical studies through pedigree charts. They are coded in MacPascal as a part of the programme for genetic research (Čundeva, 1988). Unfortunately, Mack is not widely spread and for practical use this programme has to be converted for IBM-PC compatible computer.

R E F E R E N C E S

[1] Busacker, R.G., Saaty, T.L.: Finite graphs and networks, McGraw-Hill Book Company, 1965

[2] Crow, J.F.: Beleške iz genetike, Savremena administracija, Beograd, 1967

[3] Čundeva, K.: Rešavanje problema kalkuliranja genetskih koeficienata i integrisanja genealoških grafova, Zbornik SYM-OP-IS '87, 137-144, 1987

[4] Чундева, К.: Алгоритамска поддршка за генетски истражувања, маг. работа, Универзитет „Кирил и Методиј", Скопје, 1988

[5] Elston, R.C., Stewart, J.: A General model for the Genetic Analysis of Pedigree Data, Human Heredity 21: 523-542, 1971

[6] Marinković, D., Tucić, N., Kekić, V.: Genetika, Naučna knjiga, Beograd, 1981

[7] McKusick, V.A.: Human genetics, Prentice-Hall, Englewood Cliffs, New Jersey, 1969

[8] Михајлов Д., Генетски графови, Зборник ЕТАН '87, 1987

[9] Thompson, J.S., Thompson M.W.: Genetics in Medicine, W.S. Saunders Company, Toronto, 1980

[10] Vogel, F., Motulsky, A.G.: Human genetics, Springer-Verlag, Berlin, 1979

[11] Wirth, N.: Algorithms + Data structures = Programs, Prentice-Hall, Inc., New Jersey, 1976

ГЕНЕТСКИ ДРВА И НИВНИ ГРАФИЧКИ ПРЕТСТАВУВАЊА

Катерина Чундева

Р е з и м е

Податоците за семејствата во генетските студии се сумираат и набљудуваат преку родословите. Дадени се алгоритми за нивен графички приказ со помош на сметач.