REPRESENTATIONS OF GENEALOGIES IN GRAPH THEORY: K-GRAPHS

Victor Alexandre Ploeger Mansueli
CEETEPS, Professional Master Program in Productive Systems, Brazil
victormansueli@gmail.com

MARCELO T. OKANO
CEETEPS and EAESP/FGV, Post-graduation, extension and Research Department, Brazil
marcelo.okano@cps.sp.gov.br (Corresponding)

ABSTRACT

Genealogies are studies of the ancestry of individuals, where kinships and relationships are established. Families can be viewed as social networks and, as such, their structural components are susceptible to mathematical analysis and Graph Theory, in particular, assists in diagramming genealogical representations by providing intuitive visualizations of their components and connections through three formalizations: Ore Graphs, P-Graphs and Bipartite P-Graphs. These, however, have some limitations when treated separately (siblings' relations, multiple marriages). This article proposes an alternative and complementary representation: the K-Graphs (Kinship Graphs).

Key words: Technology; Social Network Analysis; Genealogies; Graph Theory.

INTRODUCTION

How many of us are aware, in fact, of our family origins? Who among us has ever set out to search our inheritance? This article aims to aid in this quest by providing a documental form of recording our own lives, and those who have preceding us.

Preserving the history of a family is an act of reverence and respect towards our ancestors. It is a tribute for the provided legacy, which reflects directly in our present condition as links of a chain that goes back to generations. By tracing the origins of this chain, we uncover opportunities to understand contexts, situations, and circumstances that allow us to explain much of our current values and behaviours.

Searching for what ties us directly to an ancestor provides a connection with the past, seen not as an irremediable sequence of events in time, but as a guiding thread conducting to all that we are and represent today.

Genealogies, understood as studies of the ancestry of individuals, are a way to tell a story: of who we are, where we come from. They are about people, their lives, and about how their stories shape who we are (BALL, 2017).

One way to study genealogies is through social networks, objects of analysis in their structures and relationships. Social Network Analysis (SNA) is a distinct research perspective within the social and behavioural sciences; distinct because it is based on an assumption of the importance of relationships among interacting units. The concept of a network emphasizes the fact that each individual has ties to other individuals (WASSERMAN, FAUST, 1994).

Traditionally, the diagrammatic form of trees is used to visualize genealogies (a natural analogy in cases where new generations are positioned at the top, and the oldest at the base, in a "tapering" toward the common ancestors). Derivations can be found on ancestor maps where an individual appears to the left, with their ancestors on the right; or on descent maps where an individual stand at the top, in the narrowest region of an "inverted tree".

Genealogical diagrams are diagrams of graphs, for they contain individual persons as nodes, linked by relations of affinity, parenthood and siblingship (BARNES, HARARY, 1983).

In the context of genealogies and networks, Graph Theory provides three visual representations: Ore Graphs, P-Graphs and Bipartite P-Graphs (BATAGELJ, MRVAR, 2008). Each of them has its particularities, limitations, advantages and disadvantages. These characteristics are explored in this article, aiming at the proposal of a new mathematical formalization: the K-Graphs.

In this work, we have chosen to use the theoretical revision method, regarding the type of literature review. It is an eminently conceptual approach, whose synthesis of research aim at explanation building (PARÉ et al., 2015).

A theoretical review draws on existing conceptual and empirical studies to provide a context for identifying, describing, and transforming into a higher order of theoretical structure and various concepts, constructs or relationships (PARÉ et al., 2015).

Thus, this method was chosen as a means to develop a conceptual framework on the topic, where the question of initial research was refined in the proposition of a new genealogical representation.

2. Theoretical Review

2.1 Genealogies and Social Networks

Genealogy aims to establish the origin of individuals and families, through the mapping of their ancestors. Composed of a series of data, it enables the reconstruction of history by presenting, usually diagrammatically, the ancestry of an individual with the indication of successive generations. For Roso (2010), a more comprehensive definition of genealogy would be the "study of kinship."

Families are composed of people connected to each other by marriages or affiliations and, as social systems, their relational structure is constituted by the patterns of relationship between their members.

The presence of relational information is a critical and defining feature of a social network. Kinship relations have been studied using network methods for many years. Ties can be based on marriage or descent relationships and marriage or family relationships can be described using social network methods (WASSERMAN, FAUST, 1994).

Kinship is a fundamental social relation, which is extensively studied by anthropologists and historians. In contrast to people who assemble their private family trees, social scientists are primarily interested in the genealogies of entire communities (NOOY, MRVAR, BATAGELJ, 2011).

The distinctive features of kinship networks reside less in how their constitutive ties - be they biological, jural, ritual, symbolic, or whatever - are defined and established than in the way these ties are organized. Kinship networks are characterized by the interplay of three fundamental principles: filiation, marriage, and gender (HAMBERGER, HOUSEMAN, WHITE, 2011).

In historical terms, in the 1930s social networks were represented in Graph Theory, an area of mathematics whose origin dates back to 1736 (CHARTRAND, LESNIAK, ZHANG, 2010).

Diagrams of this kind have been in practical use in many parts of the world. and oral descriptions of these configurations of relations have been with us for thousands of years. Once Graph Theory was seen to have relevance for the analysis of social networks, genealogical diagrams. as graphs, became an obvious site for applying this theory (BARNES, HARARY, 1983).

No directional relations include, among other things, some kinship relations such as "is married to," or "is a blood relative of,". Graphs have been widely used in SNA as a means of formally representing social relations and quantifying important social structural properties (WASSERMAN, FAUST, 1994).

2.2 Graphs

A graph is a model for a social network with an undirected dichotomous relation; that is, a tie is either present or absent between each pair of actors (WASSERMAN, FAUST, 1994).

In its formal definition, a graph G is a finite nonempty set V of objects called vertices (the singular is vertex) together with a possibly empty set E of 2-element subsets of V called edges. Vertices are sometimes referred to as points or nodes, while edges are sometimes called lines or links (CHARTRAND, LESNIAK, ZHANG, 2010).

Figure 1 illustrates the diagram of a graph, with its vertex pairs and edges. The vertices were named after the actors involved, and the edges represent the existing relationships between each of them.

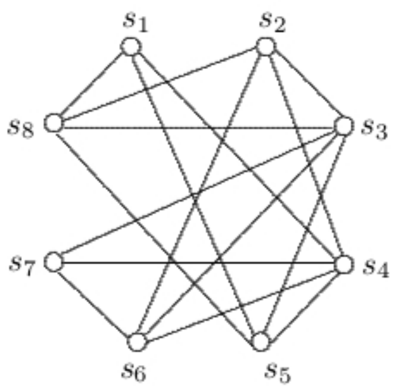

FIGURE 1: AN EXAMPLE OF A GRAPH, WITH 8 IDENTIFIED VERTICES

source: extracted from Chartrand, lesniak e zhang (2010)

2.3 Genealogical Graphs

Following are the current formalizations in Graph Theory for representations of genealogies: *Ore Graphs*, *P-Graphs* and *Bipartite P-Graphs*.

a) *Ore Graphs*

These are graphs named after the Norwegian mathematician Øystein Ore. According to Hamberger *et al.* (2011), they are the most conventional representation of a kinship network, where vertices represent individuals, arcs represent filial ties, and edges represent marriages.

In this sociogram, men are represented by triangles, women by ellipses, marriages by (double) lines, and parent-child relations by arcs. In contrast to the family tree, fathers *and* mothers are connected to their children in an Ore Graph (NOOY, MRVAR, BATAGELJ, 2011).

Figure 2 exemplifies this genealogical graph.

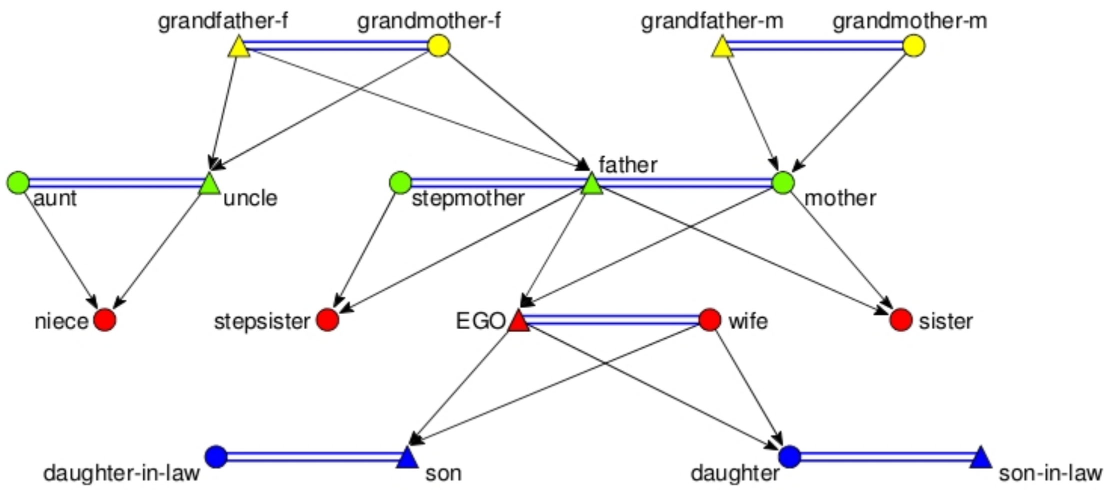

Figure 2: *An Ore Graph example*
Source: *Extracted from NOOY, MRVAR e BATAGELJ (2011)*

b) *P-Graphs*

In a *P-Graph*, couples and unmarried individuals are the vertices and arcs point from children to parents. The type of arc shows whether the descendant is male (full arc) or female (a dotted arc) (NOOY, MRVAR, BATAGELJ, 2011).

Figure 3 presents a *P-Graph* where we can observe a case of multiple marriage (father & mother and father & stepmother).

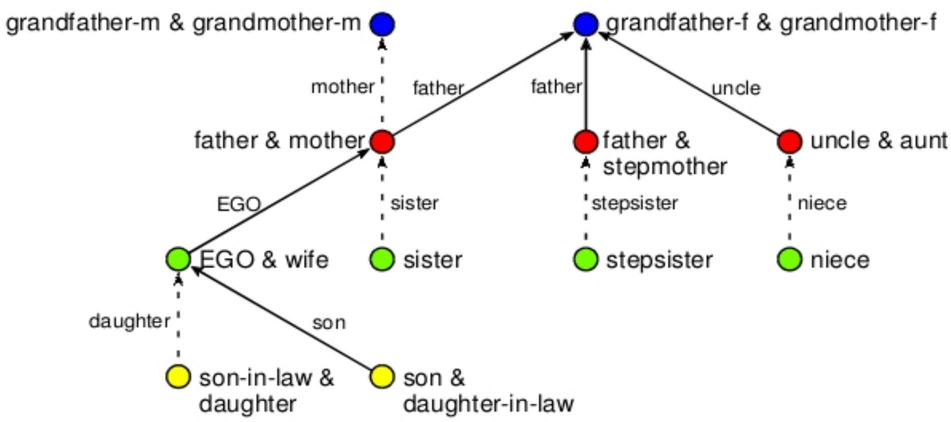

Figure 3: *A P-Graph example*
Source: *Extracted from NOOY, MRVAR e BATAGELJ (2011)*

In comparison to different presentations, Mrvar and Batagelj (2004) states that there are less vertices and lines in *P-Graphs* than in corresponding *Ore Graphs*.
Another benefit is the absence of cross-lines. However, in cases of multiple marriages, the same individual is replicated on more than one edge.

c) *Bipartite P-Graphs*
Bipartite P-Graphs has vertices for individuals and vertices for married couples. They solve the problem of distinguishing between a married uncle and a remarriage of a father or between stepsisters and nieces (NOOY, MRVAR, BATAGELJ, 2011).
White (2004) points out that half-siblings are distinguished in the Bipartite P-Graph format, in which individuals are one set of nodes and couples another. Figure 4 exemplifies this kind of genealogical graph.

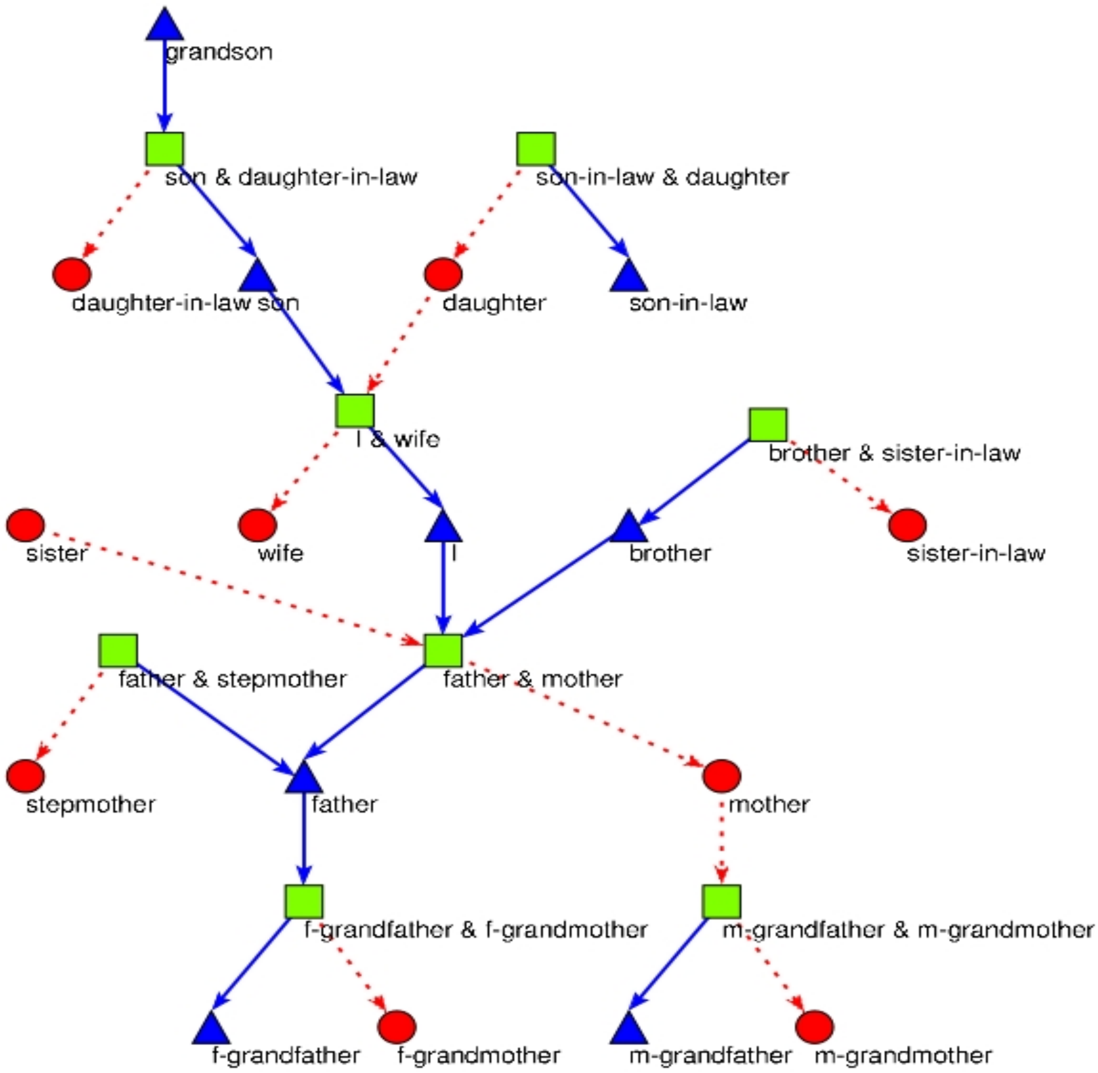

Figure 4: *A Bipartite P-Graph example*
Source: *Extracted from BATAGELJ e MRVAR (2008)*

2.4 *K-Graphs*

Conceptually, genealogical graphs are abstractions of social networks, where relationships are established between individuals with some family bond. Kinship ties are represented by agreed symbols in Graph Theory: vertices, edges and arcs (directed edges). Vertices identify individuals (or couples); edges and arcs identify relationships between individuals (in the case of *P-Graphs*, named edges are optionally used to identify affiliation (Figure 3)).
K-Graphs (K stands for "Kinship") are a new representational proposal of genealogies, which is based on the layout of the structural components of existing graphs, supplementing them by adding the main advantages of each representation. The aim was to suppress, in that way, the limitations observed when they are used alone.
The symbology and some features that distinguish *K-Graphs* from the traditional representations are listed:

- Marital relationships (pair of vertices of couples): identified by bidirectional arcs;
- Affiliations: firstborns connected to their parents by two directional arcs (from parents to child);
- Children: siblings connected together by edges (firstborns and lastborns with only one adjacent edge, "middle" siblings connected by two adjacent edges);
- Gemelarities: twins connected together by double edges;
- Genres: discriminated by the geometric figures of vertices (full triangles for male, full circles for female);
- Naming: only vertices are identified with names of individuals;
- Generations: identified by the colours of the vertices (ascending above, descending below);
- Linearity: siblings represented horizontally;
- Representation of multiple marriages of the same individual;
- Temporality: vertices that identify siblings are arranged according to the chronological order of their birth dates.

Figure 5 illustrates a *K-Graph* composed of the maternal ancestry from the first author of this article (identified by the term EGO[3] in the diagram). The components were arranged from left to right:

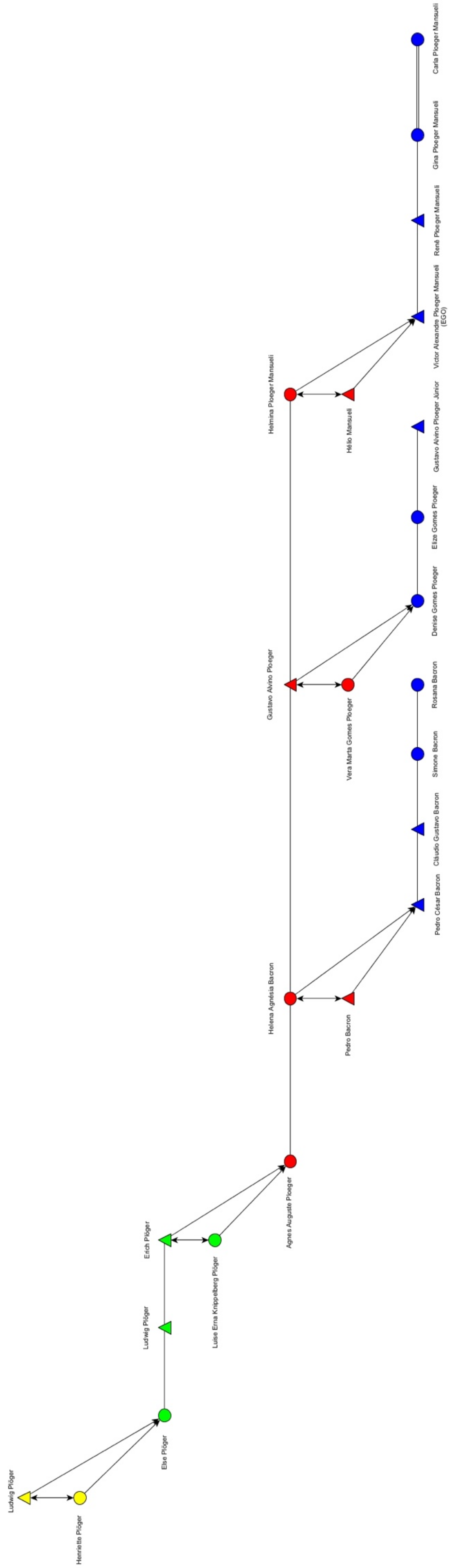
Ludwig Plöger
Henriette Plöger
Else Plöger
Ludwig Plöger
Erich Plöger
Luise Erna Knippelberg Plöger
Agnes Auguste Ploeger
Helena Agnésia Bacron
Pedro Bacron
Gustavo Alvino Ploeger
Vera Marta Gomes Ploeger
Helmina Ploeger Mansueli
Hélio Mansueli
Pedro César Bacron
Cláudio Gustavo Bacron
Simone Bacron
Rosana Bacron
Denise Gomes Ploeger
Elize Gomes Ploeger
Gustavo Alvino Ploeger Júnior
Victor Alexandre Ploeger Mansueli (EGO)
René Ploeger Mansueli
Gina Ploeger Mansueli
Carla Ploeger Mansueli

Figure 5: A K-Graph example

Figure 5 contains information on the represented family structure. Such information allows us to infer kinship relations. Following are some examples:

- 4 generations were represented, including 24 individuals (11 males and 13 females);
- The maternal great-grandparents of EGO are placed at the top of the diagram (yellow vertices);
- The maternal grandfather of EGO (lastborn) has 2 brothers (the firstborn sister and the "middle" brother);
- From his maternal side, EGO has 3 uncles (2 female and 1 male) and 7 cousins (4 females and 3 males);
- EGO is the firstborn and has 3 siblings (1 brother and 2 twin sisters);
- Of the first 3 (older) generations, 3 individuals had no descendants (Else Plöger, Ludwig Plöger (son) and Agnes Auguste Ploeger).

Table 1 compares positive and negative features of the 4 graphs presented in this study.

Table 1: Comparative table

Graph	Positive features	Negative features
Ore Graph	Direct connections between parents and children.	Absence of direct connections between siblings.
P-Graph	Direct connections between parents and children; Less vertices and lines; Absence of crossed lines.	Absence of direct connections between siblings; There's no means to distinguish half-siblings; Replication of individuals in multiple marriages.
Bipartite P-Graph	There is no replication of individuals in multiple marriages.	Absence of direct connections between siblings; More vertices and lines.
K-Graph	Direct connections between siblings; There is no replication of individuals in multiple marriages; Means to distinguish half-siblings.	Absence of direct connections between parents and children (except in the case of firstborns).

3. Results and Discussion

As can be seen from Figure 2, one of the limitations of an Ore Graph is the absence of connections between siblings' relations. In other words, it would be necessary to go back to the parents of an individual to identify other vertices originated from them, which may represent his siblings. In a similar way, go back to your grandparents (paternal and maternal), get their children and identify, among them, their uncles.

In a K-Graph (Figure 5), brothers and sisters relations are established through direct connections between siblings, with edges originating from the firstborn of a couple. Another advantage of the new representation in relation to Ore Graphs is the absence of cross-lines, which occurs when there is more than one child per couple (in a K-Graph only the first child is identified, where a directional arc originates from each of the child parents).

In a P-Graph (Figure 3), it is impossible to distinguish between a married uncle and a new marriage from a father, or between half-sisters and nieces (MRVAR, BATAGELJ, 2004). P-Graphs do not distinguish half-siblings because when the same parent is in two different marriages the half-sibling relationship resembles that of cousins (WHITE, 2004).

In addition to predicting multiple marriages (through the association, with a bidirectional arc, of a new vertex of an opposite gender to an existing vertex), K-Graphs are not prone to mistaken kinship analysis, since vertex uniqueness is guaranteed (individuals are not replicated).

Bipartite P-Graphs (Figure 4) has the drawback of containing considerably more vertices and lines than the P-Graph and path distance does not correspond to the remove of a kinship relation (NOOY, MRVAR, BATAGELJ, 2011). Furthermore, the extra components that exist in Bipartite P-Graphs impair the representation comprehensibility and readability, as the number of marriage increases.

K-Graphs identify conjugal relations through a single bidirectional arc between two vertices, therefore it is not necessary to create a new vertex to represent a couple.

CONCLUSION

This paper proposed an alternative representation of genealogies in Graph Theory: the K-Graphs. It was demonstrated that the three existing formalizations (Ore Graphs, P-Graphs and Bipartite P-Graphs) have some limitations when handled alone, and these were addressed by the new proposal: sibling relations were established through edges (which also enabled suppression of the cross-arcs observed in Ore Graphs); the possibility of representing multiple marriages was foreseen, and; a single match between individuals and vertices was assumed.

Although not addressed in this paper, K-Graphs could still represent non-traditional kinship relationships through other symbologies. Adoptions, for instance, could be identified by unfilled triangles and circles (sons and daughters, respectively).

REFERENCES

BALL, R. Visualizing Genealogy Through a Family-Centric Perspective. Information Visualization. SAGE Publications. Sage UK: London, England, 2017, v.16, p. 74-89.

BARNES, J.; HARARY, F. Graph Theory in Network Analysis. Social Networks. Elsevier, 1983, v.5, p. 235-244.

BATAGELJ, V.; MRVAR, A. Analysis of Kinship Relations with Pajek. Social Science Computer Review. Sage Publications. Sage CA: Los Angeles, CA, 2008, v.26, p. 224-246.

CHARTRAND, G.; LESNIAK, L.; ZHANG, P. Graphs & Digraphs. Sixth Edition. CRC Press, 2010.

FISCHER, M. D. Kinship Terminology. Representing Anthropological Knowledge: Calculating Kinship.

Available at: <http://era.anthropology.ac.uk/Kinship/prologTerm5.html>. Access in 17/06/2017.

HAMBERGER, K.; HOUSEMAN, M.; WHITE, D. R. Kinship Network Analysis. The Sage Handbook of Social Network Analysis, Sage Publications, 2011, p. 533-549.

MRVAR, A.; BATAGELJ, V. Relinking Marriages in Genealogies. Metodoloˇki zvezki, 2004, v.1, n.2, p. 407-418.

NOOY, W. D.; MRVAR, A.; BATAGELJ, V. Exploratory Social Network Analysis with Pajek. Cambridge University Press, 2011.

PARÉ, G.; TRUDEL, M.; JAANA, M; KITSIOU, S. Synthesizing information systems knowledge: A typology of literature reviews. Information & Management. Elsevier, 2015, v.52, n.2, p. 183-199.

ROSO, A. Psicologia e história: acerca da construção de árvores genealógicas ou como retomar lembranças de família em sociedades de rede. Psico Magazine, jul./set. 2010, v.41, n.3, p. 385-392.

WASSERMAN, S.; FAUST, K. Social Network Analysis: Methods And Applications. Cambridge University Press, 1994.

WHITE, D. R. Ring Cohesion Theory in Marriage and Social Networks. Mathematics and Social Sciences. 42e année, 2004, n.168, p. 59-82.