Genealogy-MBW

1. Introduction

Genealogy-MBW is from the Huapu System (https://www.zhonghuapu.com/) and is a real-life genealogical dataset. It begins with an ancestor (numbered 0 in the dataset) surnamed Wu who was born in Wuyuan, Huizhou in the Yuan Dynasty (in 1303), and moved to Tongcheng (now Congyang), Anqing during the Ming Dynasty. This dataset records all his descendants' data until 31 December 2020, including 23464 direct descendants with specific names. In order to facilitate sharing with genealogy culture and graph data researchers, the Huapu System provides this genealogical dataset that only contains inheritor numbers, and their private information has been anonymized.

In each genealogical graph dataset of the Huapu System, nodes represent people, edges represent relationships between people, and attributes of nodes provide the description information of people. In Fig 1, for example, 0, 1, and 7323 uniquely identify 3 nodes, where 1 and 7323 are the son of 0 and a male descendant of 16 generations apart from 0, respectively. By July 2022, the Huapu System has data records for more than 18.55 million people and 1130 genealogies.

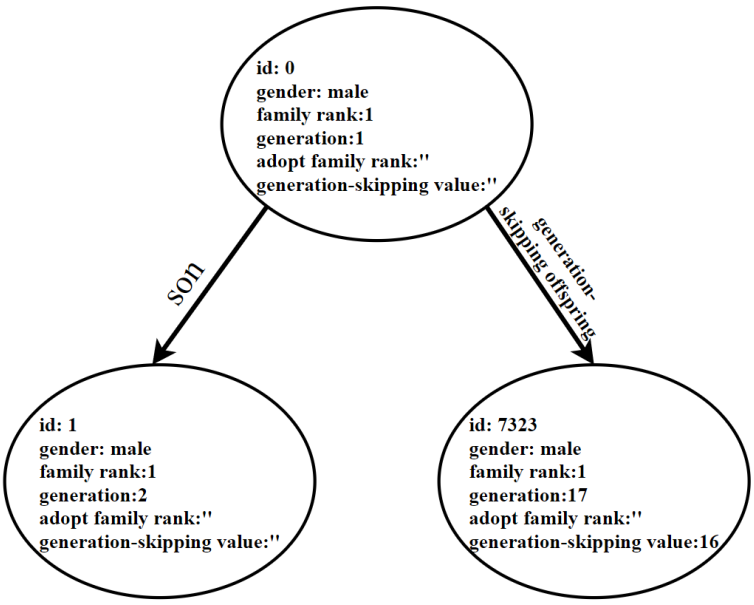

Fig 1. An example of people and relationships in Genealogy-MBW

2. Details

Table 1 provides the details of the Genealogy-MBW dataset. This dataset contains two files: Genealogy-MBW-nodes.txt provides all information about nodes, and Genealogy-MBW-edges.txt provides all information about edges. The detailed descriptions are shown in Fig 2 and Fig 3.

Table 1. Dataset description

Graph type	$\|V\|$	$\|E\|$	In/Out Degree	Number of node attributes	Number of edge attributes	Maximum number of generations
Directed	23646	24207	Min/max out degree 0/13 Min/max in degree 0/3 Min/max degree:1/14	5	10	28

Table 2 presents node attributes, attribute values, and characteristics, and Fig 2 displays nodes in the file Genealogy-MBW-nodes.txt. When a node's attribute value is empty, it means that this attribute is not applicable. For example, '0,male,2,,1,,' means a male person with an id of 0, whose family ranking is 2, his generation is 1, and his adoptive family ranking and generation-skipping values are not applicable. Similarly, '8477,male,3,1,18,,' represents a male person with an id of 8477, whose family ranking is 3, his adoptive family ranking is 1, his generation is 18, and his generation-skipping value is not applicable.

Fig 3 shows edges in the file Genealogy-MBW-edges.txt, in which the comparison table of Chinese and English relationship types is given in Table 3. In particular, there is a special relationship, generation-skipping offspring, which means that a lineal ancestor with a generation number of a person is recorded in the genealogy, while people between this person and the lineal ancestor are not available in the genealogy. To maximize the preservation of genealogy integrity, the Huapu System designs this new relationship to connect this person and his lineal ancestor. The generation-skipping value is recorded in the ancestor node.

Table 2. Introduction to node attributes

No.	Attribute (English)	Attribute (Chinese)	Attribute value	Remarks
1	id	ID	0~23643	indispensable, unique in the Genealogy-MBW
2	gender	性别	male, female	indispensable
3	family ranking	家庭排行	1~13	indispensable
4	adoptive family ranking	过继家庭排行	1~4	dispensable,
5	generation	世辈	1~28	indispensable
6	generation-skipping value	隔代相连值	2, 5, 7, 8, 9, 14, 16, 17, 19, 20, 21	dispensable

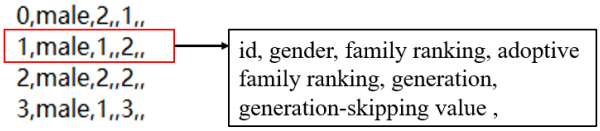

Fig 2. Genealogy-MBW-nodes.txt

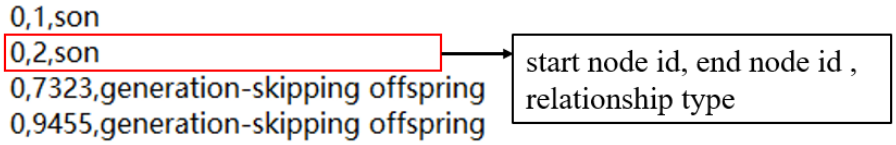

Fig 3. Genealogy-MBW-edges.txt

Table 3. Comparison table between English and Chinese relationship types

No.	Relationship type (English)	Relationship type (Chinese)
1	son	儿子
2	daughter	女儿
3	step_son	继子
4	step_daughter	继女
5	heritor	嗣子
6	foster_son	养子
7	foster_daughter	养女
8	jiantiao_son	兼祧子
9	unconfirm_offspring	待考后代
10	generation-skipping offspring	隔代后代