

SOFTWARE METAPAPER

KAMG: A Tool for Converting Blood Ties and Affinity Ties into Adjacency Matrices

Hang Xiong¹, Pin Xiong² and Hui Xiong³

¹ School of Sociology, University College Dublin, Dublin, Ireland and Department of Geography, King's College London, London, UK

² School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, China

³ School of Earth Sciences, Yangtze University, Jingzhou, China

Corresponding author: Hang Xiong (hang.xiong@kcl.ac.uk)

Kinship Adjacency Matrix Generator (KAMG) is a browser-based software for creating adjacency matrices using the information of kinship ties. Specifically, it is capable of converting the family trees in the format of GEDCOM files into adjacency matrices of blood relationship. With the data of intermarriages between different families, it can further create the adjacency matrix of affinity relationship for the families. The outcomes can be directly used to create networks. KAMG is written in JavaScript and implemented on web browsers. It is completely open source and the source code is publicly available on GitHub.

Keywords: Social Networks; Adjacency Matrix; Kinship; Family Trees; Genealogies

Funding statement: This work funded by the National Natural Sciences Foundation of China under the project "Evolution and Trigger Mechanism of Cooperative Behaviours in Rural Irrigation" (No. 71303198).

(1) Overview

Introduction

Social relationships can heavily affect individuals' decision making, and thus affect the social and economic consequences [5, 4, 6]. Kinship is one of most important social relationships. It has been studied in areas ranging from social diffusion (e.g., [12]) to immigration (e.g., [9, 11]) and employment (e.g., [1, 10]) and to social structure in general (e.g., [2, 8]).

Social network analysis (SNA) is an approach to study social relationships and the social structure they constitute using network and graph theories. It has emerged as a key technique in modern sociology, and also gained a significant following in many disciplines including economics, anthropology, biology, geography, history, and information science. Social network analysis characterises networked structures in terms of *nodes* or *vertices* (representing individuals or organisations), and *ties* or *edges* (representing connections between nodes) [7].

A typical way to mathematically represent and store a social network is using an adjacency matrix. An adjacency matrix of n vertices is the $n \times n$ matrix with a value of 1 or 0 in each element a_{ij} , corresponding to whether the node i and j are connected or not [3]. For a weighted network, the values of the elements in the adjacency matrix can be any numerical values, and they represent the weights of ties. As an example, **Figure 1** shows a weighted network (Panel (a) on the left-hand side) and its corresponding adjacency matrix (Panel (b) on the right-hand side).

Therefore, in order to conduct social network analysis on kinship ties, researchers need to represent the ties as adjacency matrices. Kinship consists blood relationship (or consanguinity), which is based on the birth, and affinity relationship, which is based on marriages. The blood relationship in a family is generally recorded as a family tree. Family trees are store digitally as GEDCOM files, plain texts (usually either ANSEL or ASCII) containing genealogical information about individuals, and the meta-data linking these records together. A detailed description of GEDCOM data format is available at the entry of "GEDCOM" in Wikipedia (<https://en.wikipedia.org/wiki/GEDCOM>). Many genealogy software (e.g., *MyHeritage Family Tree Builder*) and websites (e.g., www.familyecho.com) support creating digital family trees and exporting them to GEDCOM files. There are also genealogical forums (e.g., Roosweb www.roosweb.ancestry.com and Genealogy Forum www.genealogyforum.com/gedcom) that provide databases of GEDCOM files.

KAMG, the software we present in this paper, reads GEDCOM files as input data, calculates weights of blood ties and then creates adjacency matrices using the data of weights. Furthermore, with records of marriages between members in different families, KAMG can calculate the weight of affinity ties and create adjacency matrices of affinity relationship (including both blood ties and affinity ties) for the families involved.

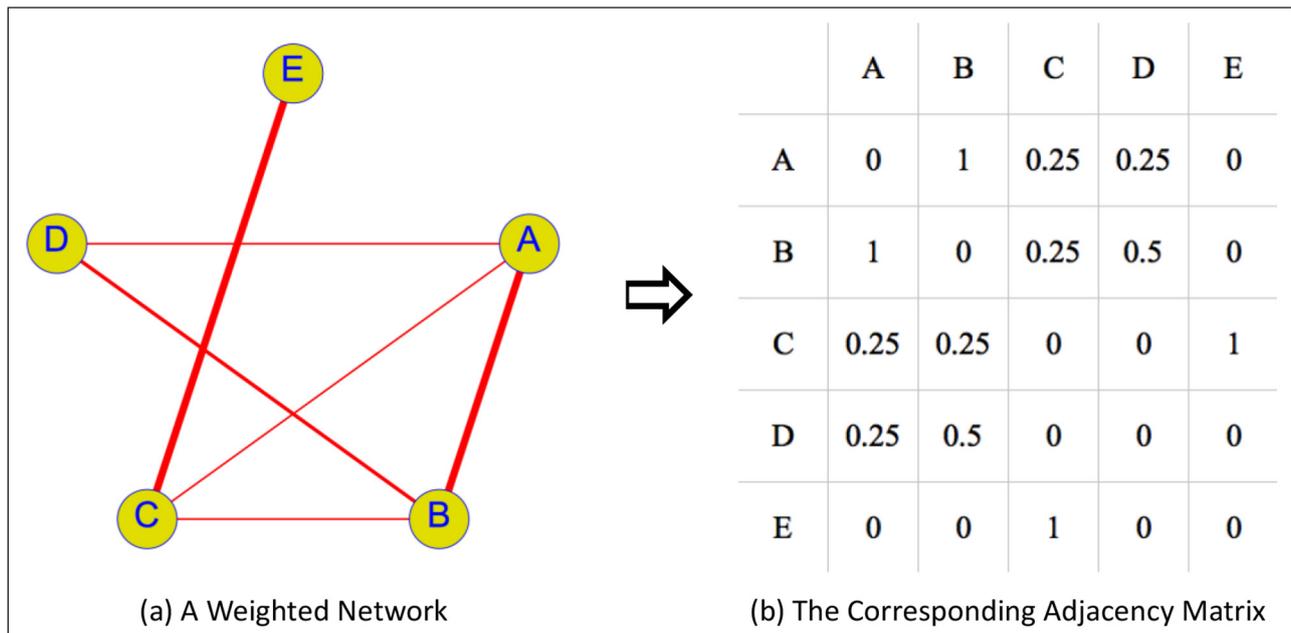


Figure 1: Network and Adjacency Matrix.

Implementation and architecture

Figure 2 displays the steps of generating adjacency matrices of blood relationship (Step 1 to 4) and an adjacency matrix of affinity relationship (Step 5 to 7) using KAMG.

The steps to generate an adjacency matrix of blood relationship using KAMG are as follows.

First, read the addresses of families. Each family has its own address (different families can share an address). This

function is helpful when researchers want to differentiate families by their locations. In cases where this differentiation is not needed, users can assign a same address to all families, but it is required to assign an address for each family. The address in the software is set to have two levels: Level 1 and Level 2, indicating a larger and a smaller address respectively (e.g., town and village). Address records need to be manually created by users. In

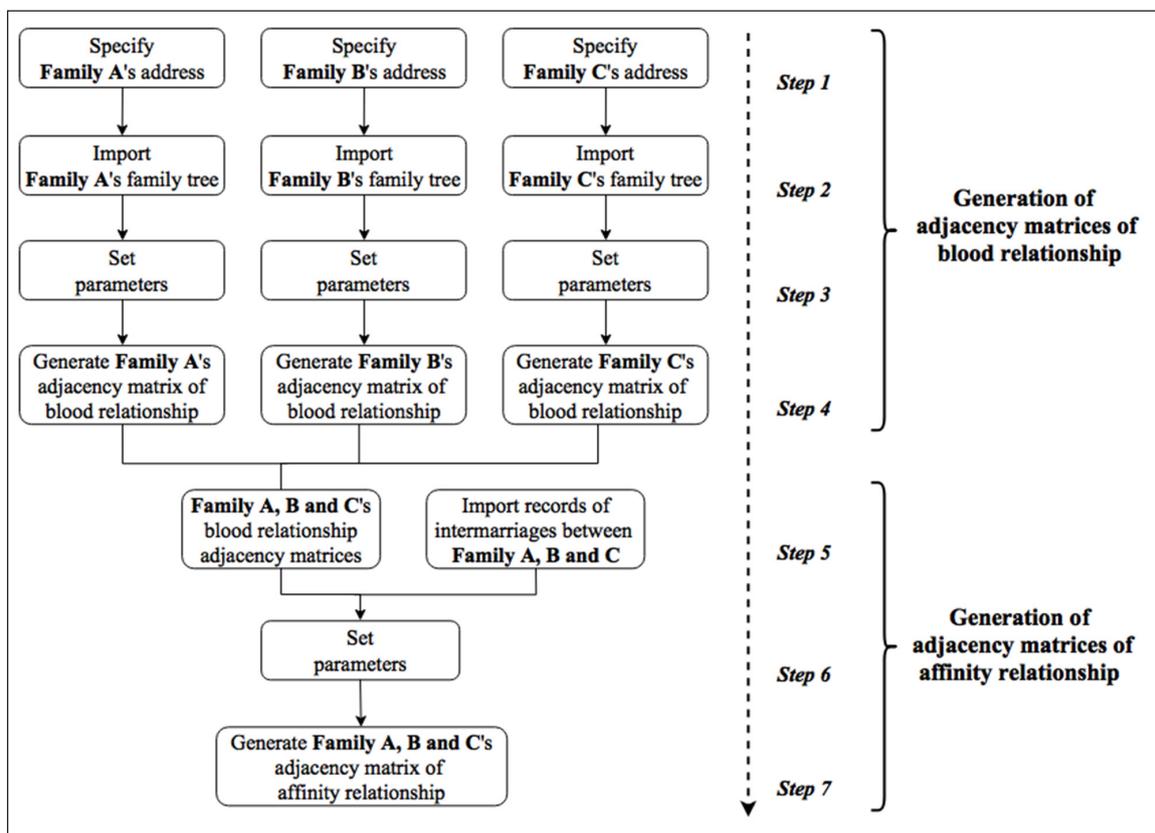


Figure 2: Steps of Creating Kinship Adjacency Matrices using KAMG.

the present version, they should be stored in a text file (maybe be extended to other file format, such as CSV file, in future versions).

Second, import family trees in GEDCOM files. As long as addresses have been successfully read (a pop-up hint window will appear), the families under the addresses will appear in a drop-down list in the “Select Clan and Members” area. Users can then select the family for which the adjacency matrix of blood relationship is planned to create from the drop-down list. Next, one can import the family tree (in a GEDCOM file) of the selected family. As long as the family tree is successfully imported (a pop-up hint window will appear), the members in the family will be listed in the “Members in the family are displayed below” area. Users can then select the members they want to include in the adjacency matrix. Finally, users need to store the selected members by applying the “Save” button; otherwise no member will be selected for the family. Repeat this process for each family tree till all family trees are imported.

Third, set the parameters for calculating weights of blood ties, i.e., algorithm and decay coefficient of consanguinity. The meaning of these parameters will be elaborated in Section “Calculation of weights of relation”. The software memorises the latest setting.

Fourth, generate adjacency matrix of blood relationship. This is simply to apply the “Generate” button. Then, an adjacency matrix with all the stored members and the weights of the blood ties between one and another will be generated, and stored in a CSV file (a pop-up window indicating its directory path will appear).

Having adjacency matrices of blood relationship for multiple families, users can further use KAMG to create an adjacency matrix of affinity relationship if the data of intermarriages between these families are available. Note that, to be consistent, the weights of blood relationship for these families should be calculated under the same parameters. The steps of creating an adjacency matrix of affinity relationship are as follows.

Fifth, import the intermarriage record. This is a record of the information of intermarriages between members in different families. A record should contain the names of the two members involved and their families (along with their addresses). Intermarriage records need to be manually created by users. In the present version, they should be stored in a text file (maybe be extended to other file format, such as CSV file, in future versions).

Sixth, set the parameter for calculating weights of affinity ties, i.e., decay coefficient of affinity. The meaning of the parameter will be elaborated in Section “Calculation of weights of relation”.

Seventh, generate adjacency matrix of affinity relationship by applying the “Generate” button, like in the fourth step. A large adjacency matrix including weights of both blood ties and affinity ties for all members in all the families involved will be generated, and stored in a CSV file (a pop-up window indicating its directory path will appear).

A video demonstrating how to use KAMG is publicly available at <https://www.youtube.com/watch?v=FlcinderOHY>.

Calculation of weights of kinship ties

The fourth step and the seventh step involve the calculation of weights of kinship ties. We need to decide the weights of blood ties first, and then compute the weights of affinity relations based on the weights of blood relations.

Weights of blood ties

The weight of a blood tie between two members is dependent on how close in blood the two members are; that is, the fraction of genes they share. The blood proximity is measured by the *degree of consanguinity*. Two algorithms are widely used to compute the degree of consanguinity: The algorithm based on the Canon Law (a body of laws and regulations made by ecclesiastical authority for the government of a Christian organization or church and its members) and the algorithm based on the Roman Law (a legal system applied in most of Western Europe until the end of the 18th century). The algorithms are conducted in the following steps:

1. Trace back to the nearest common linear ancestor of these two members.
2. Count how many generations between the common ancestor and the two members respectively (two number will be obtained accordingly).
3. If following the Canon Law, take the number with larger value as the degree of consanguinity. If following the Roman Law, take the sum of the two numbers as the degree of consanguinity.

Suppose the degree between two members is b , the weight of blood relation between them is defined as

$$w_b = d_b^{b-1}$$

where d_b is the decay coefficient of consanguinity, a parameter indicating how much the weight decays as the degree of consanguinity increases. The value is generally set as 0.5.

Taking the Canon Law algorithm as an example, the degree of consanguinity between one and his parents or siblings or children is 1. The weight of the blood tie is hence $0.5^{1-1} = 1$. This is the highest possible weight for blood ties. The second highest weight is $0.5^{2-1} = 0.5$, which measures the weight between one and his grandparents, uncle and aunt, niece and nephew, or grandchildren.

In **Figure 3**, for instance, the degree of consanguinity between $D01$ and $C03$ in the Payne family (on the left-hand side) is calculated as follows. First, find their nearest common linear ancestor A , who is $D01$'s great grandfather and $C03$'s grandfather. Second, count the number of generations between A and $D01$ and between A and $C03$, which are 3 and 2 respectively. Third, take the larger number, 3, as the degree of consanguinity. The consanguinity weight $D01$ between $C03$ and is thus $0.5^{3-1} = 0.125$.

Weight of Affinity Relations

As mentioned earlier, the marriage between two members in different families creates an affinity relation. The weight of the affinity relation between the individual i in the

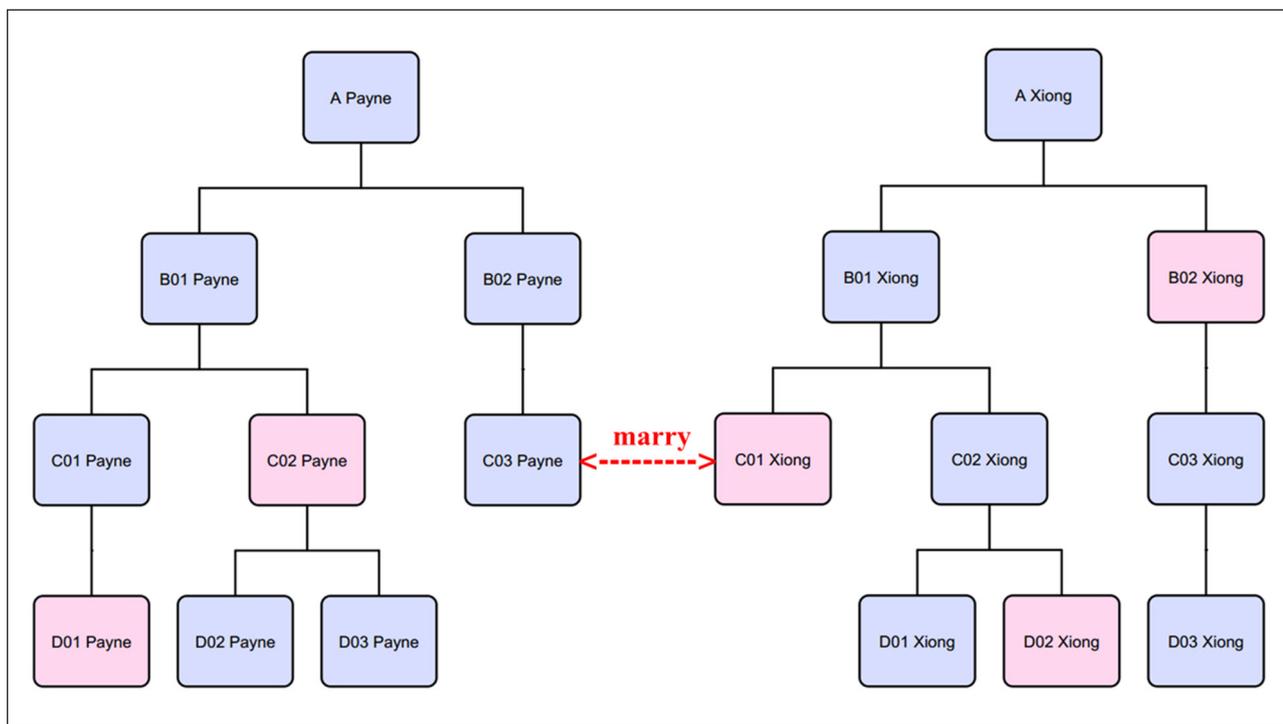


Figure 3: Family Trees.

husband's family and the individual j in the wife's family $w_a(i, j)$ is given by

$$w_a(i, j) = d_a w_b(i, m) w_b(j, f)$$

where $w_b(i, m)$ is the weight of the blood tie between i and the husband, $w_b(j, f)$ is the weight of the blood tie between j and the wife. And d_a , like d_b , is a parameter indicating how much the weight decays because this relation is inter-family. It is thus named the decay coefficient of affinity, and its value is set as 0.5. The weight of the tie between, say, the husband's uncle and the wife's father is $0.5 \times (0.5 \times 1) = 0.25$.

In **Figure 3**, suppose the man $C03 Payne$ marries the woman $C01 Xiong$. The weight of the affinity tie between the man's second-degree brother $C01 Payne$ and the woman's second-degree nephew $D01 Xiong$ is calculated in two steps. First, measure the weight of the blood tie between $C01 Payne$ and the husband, which is 0.5, and the weight of the blood tie between $D01 Xiong$ and the wife, which is 0.25. Second, apply the above formula. We obtain $0.5 \times (0.5 \times 0.25) = 0.0625$.

Quality control

The software conducts two calculations: calculation of weights of blood ties and calculation of weights of affinity ties. Both are not complicated and the results can be checked by hand. We validated the calculation results in two ways. One is checking by hand and the other is comparing the results with those calculated using equivalent R codes (the algorithms used in the software is originally coded in R). The validations were carried out with both a small number of families and intermarriages, and a large number of families (more than five, and each with a few hundreds of members) and intermarriages (more than thirty).

The software has been tested both experimentally and in practice. First of all, each of the three authors independently created a number of family trees and intermarriages and used them as input data to test how the software ran. They carried out the tests on different web browsers Internet Explorer 6 up to 10. Furthermore, the first author used the software to conduct a study the network structure of villages consisting more than 400 households split in 64 families.

(2) Availability

KAMG is open source software. It has been published on GitHub under the GNU General Public License.

Operating system

Windows XP and higher

Programming language

JavaScript

Additional system requirements

None.

Dependencies

This software needs to be run on a web browser, and it relies on the ActiveX plugin of web browser for importing and exporting data. It therefore can only function properly on web browsers that allow the ActiveX plugin to run, typically Internet Explorer 6 to 10.

List of contributors

Hang Xiong (School of Sociology, University College Dublin; Department of Geography, King's College London)
 Pin Xiong (School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology)

Hui Xiong (School of Earth Sciences, Yangtze University)

Software location

<https://github.com/hangxiong/KAMG>

Language

English

(3) Reuse potential

The algorithms of calculate weights of affinity ties can be used to calculate the weights of any kind of connections between two members in different networks. This process, in a circumstance of multiple networks, involves the selection of weights that are contradicted for different ways in which a connection is created, and for different time sequence in which a connection is created.

So far, the software can only be implemented on Internet Explorer web browsers. Users can extend it to be compatible with other types of browsers (such as Google Chrome, Mozilla Firefox).

Acknowledgements

The authors thank Prof. Diane Payne at School of Sociology, University College Dublin for her support in developing the software tool.

Competing Interests

The authors declare that they have no competing interests.

References

1. **Assaad, R** 1997 Kinship ties, social networks, and segmented labor markets: evidence from the construction sector in Egypt. *Journal of Development Economics*, 52(1), pp. 1–30. DOI: [http://dx.doi.org/10.1016/S0304-3878\(96\)00441-5](http://dx.doi.org/10.1016/S0304-3878(96)00441-5)
2. **Curry, O, Roberts, S G B and Dunbar, R I M** 2013 Altruism in social networks: Evidence for a “kinship premium.” *British Journal of Psychology*, 104(2), pp. 283–295. DOI: <http://dx.doi.org/10.1111/j.2044-8295.2012.02119.x>
3. **Diestel, R** 2006 *Graph Theory*. Available at: <http://books.google.com/books?id=aR2TMYQr2CMC&pgis=1>.
4. **Granovetter, M** 1985 Economic action and social structure: the problem of embeddedness. *American journal of sociology*. Available at: <http://www.jstor.org/stable/10.2307/2780199> [Accessed October 23, 2013].
5. **Hinde, R A** 1976 Interactions, Relationships and Social Structure. *Man*, 11(1), pp. 1–17. Available at: <http://www.jstor.org/stable/2800384>. DOI: <http://dx.doi.org/10.2307/2800384>
6. **Jackson, M O** 2014 Networks in the Understanding of Economic Behaviors. *Journal of Economic Perspectives*, 28(4), pp. 3–22. Available at: <https://www.aeaweb.org/articles.php?doi=10.1257/jep.28.4.3> [Accessed September 4, 2015]. DOI: <http://dx.doi.org/10.1257/jep.28.4.3>
7. **Jackson, M O** 2008 *Social and Economic Networks*. Princeton University Press.
8. **Janečka, J E** et al 2006 Kinship and social structure of bobcats (*Lynx rufus*) inferred from microsatellite and radio-telemetry data. *Journal of Zoology*, 269(4), pp. 494–501. DOI: <http://dx.doi.org/10.1111/j.1469-7998.2006.00099.x>
9. **Menjivar, C** 1995 Kinship Networks Among Immigrants: Lessons From a Qualitative Comparative Approach. *International Journal of Comparative Sociology*, 36(3/4), pp. 219–232. DOI: <http://dx.doi.org/10.1177/002071529503600307>
10. **Montgomery, J D** 1991 Social Networks and Labor-Market Outcomes: Toward an Economic Analysis. *The American Economic Review*, 81(5), pp. 1408–1418.
11. **Shaw, A** 2001 Kinship, Cultural Preference and Immigration: Consanguineous Marriage Among British Pakistanis. *Journal of the Royal Anthropological Institute*, 7, p.315. Available at: <http://dx.doi.org/10.1111/1467-9655.00065>
12. **Xiong, H and Payne, D** 2015 *Characteristics of Chinese Rural Network: Evidence from Villages in Central China*. Available at: <http://ssrn.com/abstract=2621621>.

How to cite this article: Xiong, H, Xiong, P and Xiong, H 2016 KAMG: A Tool for Converting Blood Ties and Affinity Ties into Adjacency Matrices. *Journal of Open Research Software*, 4: e28, DOI: <http://dx.doi.org/10.5334/jors.99>

Submitted: 08 October 2015 **Accepted:** 01 July 2016 **Published:** 19 July 2016

Copyright: © 2016 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.