Journal of Computational Social Science
https://doi.org/10.1007/s42001-023-00245-4

RESEARCH ARTICLE

Intra-family links in the analysis of marital networks

J. J. Merelo[1,2] · **M. Cristina Molinari**[3]

Received: 25 July 2023 / Accepted: 17 December 2023
© The Author(s) 2024

Abstract
Marriage networks, which represent the matrimonial connections between different families in a given historical and geographical milieu, rarely take into account one aspect of internal family dynamics, namely the existence of intra-family marriages. The inclusion of such marriages, represented in the graph by self-loops, is essential to compute more accurate measures of centrality. In this paper, we discuss various procedures for incorporating these links into the analysis, with the requirement that they be compatible with the use of already available social network analysis software. We then apply them to two historical marriage networks, one from the Republic of Venice and the other from Taiwan. By comparing centrality measures for the baseline and modified networks, we found that the most satisfactory of the proposed methods is the one that duplicate nodes of families with intra-family marriages and adds new edges that link these duplicated nodes to all the families to which the original node was connected. This procedure is computationally simple and conceptually sound, making it a useful tool for analyzing marital networks.

Keywords Social network analysis · Alliance networks · Self-loops · Marital networks · Pseudographs · Centrality measures

✉ J. J. Merelo
jmerelo@ugr.es
https://scholar.google.com/citations?user=gFxqc64AAAAJ

M. Cristina Molinari
cmolinar@unive.it
https://scholar.google.com/citations?user=MW116skAAAAJ

[1] Department of Computer Engineering, Automatics and Robotics, University of Granada, Granada, Spain

[2] CITIC, University of Granada, Granada, Spain

[3] Dipartimento di Economia, Ca' Foscari University of Venice, Venice, Italy

Published online: 25 January 2024

Introduction

Social network analysis has been widely used to study marital alliances in a given polity or historical milieu and to obtain insights on economic, political, and sociological issues.[1] Marriage (or marital) networks are usually built on a family level and they are represented by a graph where the nodes (vertices) are families, and an edge (arc) linking two nodes indicates the number of marriages between members of the two families (weighted network) or, less often, whether there is at least one matrimonial link or none between them (unweighted). More precisely, starting from a set of individuals and their matrimonial ties, the former is partitioned into families along patriarchal line, so a family is a group of relatives with a common surname. This is the usual simplification to an unipartite network of what is, to start with, a bipartite graph where the two "parties" are the female and the male partitions of the family, and an edge joins the node of the groom's family to the one of the bride's family. A consequence of this simplification is that an intra-family marriage, for example a marriage between two cousins with the same last name, is represented by an edge between two separates nodes in the bipartite graph but is transformed into a self-loop, i.e., an edge that originates from and terminates in the same node, because in the unipartite network, the groom's and the bride's family nodes are the same.

The emergence of self-loops has disadvantages: since self-loops are often dropped in the analysis, useful information about the network of alliances is lost. Clearly, the loss in information depends on the number of self-loops which, in turn, is linked to the size of the families and their willingness to arrange internal marriages and still be considered a single family. Moreover, the existence of self-loops is also a cultural phenomenon. For example, in much of Africa marriage is exogamic and used as a way to create bonds between different clans, whereas in South Asian countries or in the Jew tradition, endogamic marriage is socially accepted and even encouraged [1, 2].

In this paper, we are interested in how self-loops can affect centrality measures [3], such as betweenness centrality, eigenvalue centrality, and PageRank. Centrality measures fall roughly in two fields: those using the adjacency matrix, like eigenvector centrality, and those using paths over the network. Since self-loops are represented by non-zero values on the diagonal of the adjacency matrix, the first type of measures could take self-loops into account and use them to compute matrix eigenvalues (eigenvector centrality) [4]. On the other hand, path-based measures typically work by *traveling* or measuring the connections between one vertex and another, and such paths generally exclude self-loops. However, even when self-loops could be used in the analysis, they are usually dropped before creating the graph from which the computations are made, so they are not taken into account in any meaningful way.

[1] For a list of contributions, see "A survey of marriage networks".

Technically, networks that include self-loops are pseudographs or multigraphs [5],[2] whose main difference with regular graphs is the fact that there can be several edges joining two nodes or, as in this case, a node with itself[3]; this is why, it could be argued that an effective treatment of networks that include self-loops should take this fact into account. Unfortunately, and despite pioneering efforts [7] and [8], for the time being, it is impossible to perform comprehensive social network analysis, including node-level measurements as well as meso-level structures (modular structures such as communities), if marital networks are represented using multigraphs.[4]

Due to the lack of either social analysis algorithms or software capable of handling self-loops, most recent papers [9, 10] keep applying the usual methodology of collapsing these graphs to single-mode graphs and then employ off-the-shelf social network analysis tools (such as the `igraph` R or Python library [11] or *Gephi* [12]). The use of such tools is essential for the democratization of social network analysis in historical, anthropological, and sociological contexts, and being able to use them is set as the foundation of the proposals we present in this paper.

To study the influence of self-loops in the analysis of marital networks, we propose a selection of methods for converting bipartite networks that include self-loops into single-mode networks. Before presenting our proposals, we note that since centrality measures use paths over edges for their computation, any method to incorporate self-loops involves the creation of new nodes and new edges connecting these "artificial" nodes to the old ones. Therefore, our alternative methods will differ in the way nodes and edges representing intra-marriages are incorporated into the graph.

After describing three alternative methods for incorporating self-loops, we apply them to a dataset of marriages from the Republic of Venice and discuss their relative merits. Note that it is complicated or even impossible to use an external measure to prove that the centrality values obtained by including self-loops are better than those obtained by excluding them. Therefore, we will validate our method empirically, first by testing whether it is able to meaningfully incorporate self-loops into the calculation of centrality measures, without introducing uninterpretable artifacts that would obscure the analysis; and, second, whether the resulting ranking of nodes is able to better represent the status and position of families, as described in the original article in which the dataset was first introduced.

Once we have chosen the method that best satisfies the above requirements, we apply it to a second marriage network of Taiwanese elite families; this network has very different characteristic in terms of both the number of intra-marriages to be included, and the structure of the graph. This will allow us to check the robustness of our method.

[2] Some authors [6] distinguish between pseudographs (multiple edges between nodes *and* loops) and multigraphs (multiple edges between nodes, but no loops); however, nowadays, there is usually no distinction between them. Please note that there is no specific name for graphs with loops but *without* multiple edges between nodes.

[3] Admittedly, this is a degenerate form of multigraph, since it only has double edges in the case a node is connected to itself, and just a single (weighted) edge.

[4] As a matter of fact, a marital network can be embedded in a larger network that might include other types of kin connections or even commercial or political relationships; so it is indeed a multigraph.

This paper focuses on marital networks, but the importance of self-loops is also acknowledged in other contexts. For example, [13] proposes a method to analyze private vehicle commuting traffic networks in cases where intra-county traffic connections are significant. Their model, called CCME-SL, is able to account for self-loops in community detection algorithms; [14] and [15], instead, discuss the importance of self-connections between nodes when studying the persistence of metapopulations in geo-ecological networks and suggest new network metrics to account for them. Different contexts have different interpretation on self-loops: in an online social network context [16], it refers to re-posts of former content, for instance, and the paper studies its influence in the context of information diffusion among support groups. In an epidemiological setting [17]; it would be equivalent to self-contagions or contagions among members of the same community, although the cited paper analyzes commercial networks and its influence in the spreading of swine fever. In general, providing a tool that is able to use self-loops beyond high-level measures (like its number or existence in certain nodes) will contribute to a deeper understanding of social network dynamics in many different contexts.

The rest of the paper is organized as follows: the next section is a brief survey of analyses of marital networks; next, "Datasets" presents the datasets we will be using; "Methods and results" describes the steps of each proposed method and applies them to a specially selected social network, the Venetian Republic marital network; we then validate the methods that meets our requirements in a second network, the Taiwanese elite families network. A brief discussion follows in "Discussion". Finally, our conclusions are presented in "Conclusion".

This paper has been developed in an open science environment, following the principles of Agile Science [18]. This guarantees in-time delivery, as well as a clear problem-solving orientation from the beginning. Milestones in the development of the paper can be checked in its repository.

A survey of marriage networks

Marriage and kinship networks are an interesting source of insights into the social, economic, and political dynamics of polities below a certain size, where families are strongly linked by mechanisms of economic or social inheritance. After Padgett and Ansell's pioneering analysis of marriage networks in the Grand Duchy of Florence [19], they have been explored in many different cultures and historical periods: marriages in medieval Venice have been studied to shed light on the pattern of long-distance trade [20] and on the careers of politicians [21]. In the Republic of Venice, access to power was restricted to aristocrats and nobility was hereditary. This is why, these two papers include all available marriages between noble families; [22], on the other hand, limits his attention to the families of doges, the heads of the republic. Marriages in the Venetian Republic territory of Ragusa (present-day Dubrovnik) in the sixteenth, eighteenth, and nineteenth centuries are the focus of [23].

Marital networks in East Asian countries have also been extensively investigated: [9] examines Taiwanese elite families (1895–1996), [24] the Joseon Dynasty

in Korea (1476–1910), and [25] the Tang aristocracy in China (618–906). Moving to Southeast Asia, namely the Philippines, [26] uses family network centrality to explain mayoral elections results in the first decade of the 2000s and [27] analyzes the family networks of bureaucrats and their relationship with the effectiveness of public service delivery. To conclude this long list, Haitian elites are the subject of [28], and [10] examine marriages among 'Ndrangheta families in the South of Italy. For more references, see also [29].

Centrality measures are at the heart of most of the papers listed above, but self-loops are not. This article attempts to make a contribution in this area by discussing different ways of including intra-family connections and empirically testing them to find the most appropriate. We will do this using two datasets that apply the usual approach of turning the bipartite marriage network to a unipartite network and that include a significant number of intra-family connections overlooked in the original study. They are the Republic of Venice network by Puga and Trefler [20] and the Taiwanese elite family network by Dluhošová [9] that we present in the next section.

Datasets

The dataset of marriages involving a noble husband in the Republic of Venice, from 1348 to 1887,[5] is based on records from the *Archivio di Stato di Venezia* and was digitized by Puga and Treffler [20]. In the process, family names were normalized to the most common spelling.[6] We have also eliminated all marriages that include a non-patrician wife.[7] The unipartite, undirected, and weighted network thus obtained includes 348 nodes and 12227 arcs. The total number of intra-marriages in this network is 385, 3.15% of all marriages.

The next dataset we are going to use is Dluhošová's marital network of Taiwanese elite families [9]. The original database includes several types of kin relationships, from which we extracted the marriages to obtain a undirected weighted unipartite network as before. The original dataset includes family names in Chinese characters. We processed these names using machine translation and some manual corrections, so that family names in this paper match those in the original one. In the resulting network, there are 1243 nodes and 1365 edges, making it much more sparse than the previous one. Out of the total number of edges, only 18 are self-loops, that is, 1.32% of the total number of marriages. This percentage is lower than the 3.15% of the Venetian network, but not in a totally different order of magnitude.

[5] The republic fell in 1797, but marriage records are available until 1887.

[6] Venetian family names sometimes have different spellings in the records, alternating between Venetian and Italian spelling. For instance, "Cornaro" and "Corner" have been normalized to "Corner".

[7] In the original dataset, these were marriages where the family name of the wife was not available, and therefore, they could not be properly assigned to a node in the social network. These records can be used to study which families were more prone to marrying non-patricians and when, but they are not necessary for the purposes of this paper.

Table 1 Summary of the original datasets used in this paper, which shows the differences between them in almost all aspects

Dataset	Venice	Taiwan
Main component size	99.43	92.12
Marriages	12227	1365
Self-loops (% of all marriages)	3.15	1.32
# Families	348	1243
# Families with self-loops	79	16
% Families with self-loops	22.70	1.29
Degree centrality (avg)	70.270115	2.196299
Closeness centrality (avg)	0.007006965	0.012951374
Clustering coefficient (avg)	0.449487556	0.006434346

A summary of the two datasets is shown in Table 1; one can see that they are quite different from most points of view, the main one being the clustering coefficient. Both, however, correspond to cultures where the concept of "family" spans several generations, and, most importantly, include self-loops, so they are both adequate for our purposes. Please check the Declarations section for data and code availability.

Methods and results

In this section, we propose three different methods to include self-loops in the analysis of marriage networks, and apply them to the Venetian dataset; as this network has the largest number of self-loops, it can give us a better idea of the usefulness of our proposals. The results obtained on this network will allow us to choose what looks the most appropriate method among the three proposed. To verify the quality of our selected method, we apply it again to the Taiwanese dataset, a marital network very different from the Venetian one.

For each method, we will first check whether it is able to meaningfully incorporate self-loops into the calculation of centrality measures, and second, whether the resulting ranking of nodes is able to better represent the status and position of families described in the original papers where the datasets were first introduced.

We will consider three of the measures most commonly used in social network analysis: betweenness, eigenvector, and PageRank centrality. Betweenness centrality [30] is a measure of brokerage and bridging, that is, of how well one family is able to intermediate between the others; it is a good first approximation of a family's power, reputation, or influence. Eigenvector centrality [31, 32] has been used extensively in social network analysis and takes into account not only how well connected is a node but also the importance of such connections; PageRank centrality [33], as EV centrality, is defined recursively, but is based on the importance of all the in-coming ties. Other centrality measures would either be unrelated or unaffected by self-loops, such as closeness centrality, or would be affected in a trivial way, for example degree centrality.

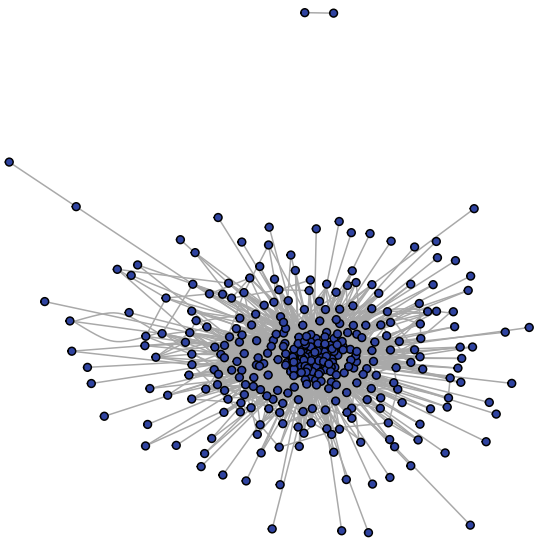


Fig. 1 Venetian marital networks with self-loops eliminated; this will be our baseline graph

Table 2 Top ten families in the Venetian dataset, with self-loops excluded, according to the three centrality measures: betweenness (left), EV centrality (middle), and PageRank (right)

No self-loops					
Family	Betweenness	Family	EV centrality	Family	PageRank
Contarini	7567.184	Contarini	1.0000000	Contarini	0.0408753
Corner	4788.719	Morosini	0.8175570	Morosini	0.0277745
Querini	4618.334	Corner	0.6425724	Corner	0.0268341
Donato	3796.409	Querini	0.5045249	Querini	0.0236407
Morosini	2864.487	Priuli	0.4286878	Donato	0.0189244
Dolfin	2398.849	Dolfin	0.3771834	Priuli	0.0183038
Balbi	2273.476	Giustinian	0.3719008	Malipiero	0.0169405
Michiel	2231.252	Michiel	0.3692852	Michiel	0.0160372
Priuli	2164.769	Zorzi	0.3629267	Loredan	0.0156105
Malipiero	1993.982	Loredan	0.3593036	Zorzi	0.0153477

The baseline network we will be working with is shown in Fig. 1; self-loops have obviously been dropped. For the following discussion, values of centrality measures calculated on this dataset will be used as a benchmark. The rankings of the top ten Venetian families for the three centrality measures we consider in this paper are shown in Table 2.

Instead, Table 3 shows the top ten families in terms of the number of intra-marriages. As we can see form the table, the Contarini family has the

Table 3 Top ten families by number of intra-family marriages

Family	# Intra-family marriages
Contarini	65
Morosini	23
Corner	19
Martinengo	16
Querini	15
Balbi	12
Donato	11
Malipiero	10
Zorzi	10
Zancaruol	9

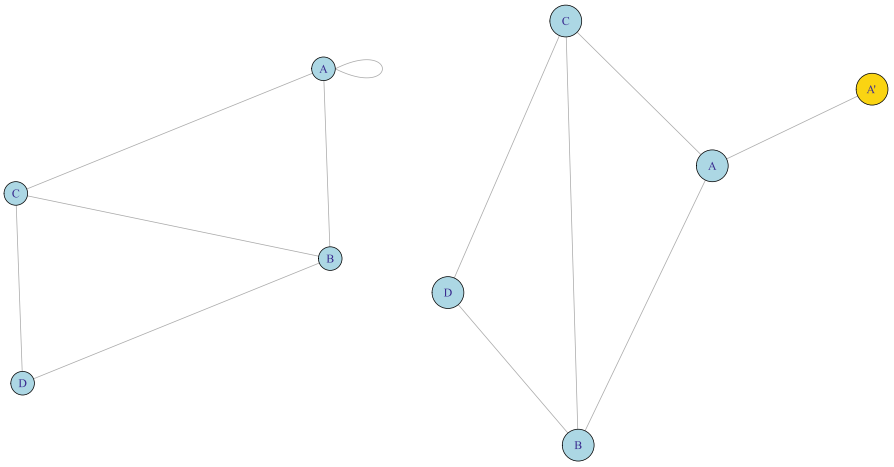


Fig. 2 Illustration of the "new node" method, with the original network with a self-loop on the left and the modified network on the right. The original nodes are light blue; the "replica" ones gold

highest number of internal marriages, accounting for 5% of the total number of marriages.

Before presenting our proposals, notice that to incorporate self-loops, any method must create new nodes and new edges linking the new, "artificial" nodes to the old ones, since methods such as the PageRank and betweenness centrality use paths over edges for their computation; new edges will have to match, somehow, the intra-links we want to incorporate into the graph modeling the social network. Therefore, the proposed method will differ in how they create new nodes.

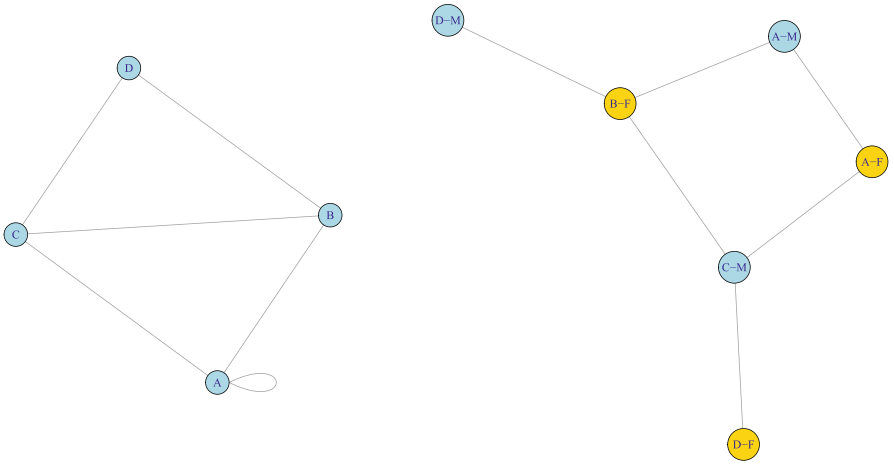


Fig. 3 Illustration of the "split families" method, with the original network with a self-loop on the left and the modified network (via the "split families" method) on the right. Female nodes have been painted gold. We have assumed that there is a single marriage between nodes, and thus, C and B need to be either male or female each, while D is split in two different (and unconnected) nodes, since they do not have a self-loop

Method 1: "New nodes"

The first method tested is relatively simple and straightforward: for each node with n self-loops, add a new node connected *only* to the original one, with a weight equivalent to n. This is illustrated in Fig. 2. In other words, we convert a self-loop into an edge between two nodes for the same family: the original node and the new one.

This method, trivial as it is, does not change the overall shape of the network, so the structural influence of these "new" nodes (and their edges) is lost. Once rendered, it might help visualize the placement of the families with some degree of intra-marriages, but little else; this could be achieved in other ways that do not involve changes in the network, such as the use of size or color in the visualization of nodes. Thus, we discard this method altogether.

Method 2: "Split families"

A different approach for taking intra-family ties into account is to consider husband and wives as different vertices of the graph; this would convert the "raw" bipartite graph (with the two parties being "bride" and "groom" nodes) in a single-party graph by simply relabeling the graph as a single-mode graph and analyzing it as such; this is illustrated in Fig. 3. From a historical perspective, this makes sense only in contexts where marriages are not egalitarian, and female and male parts of a family are separate actors, belonging to different "classes"; but,

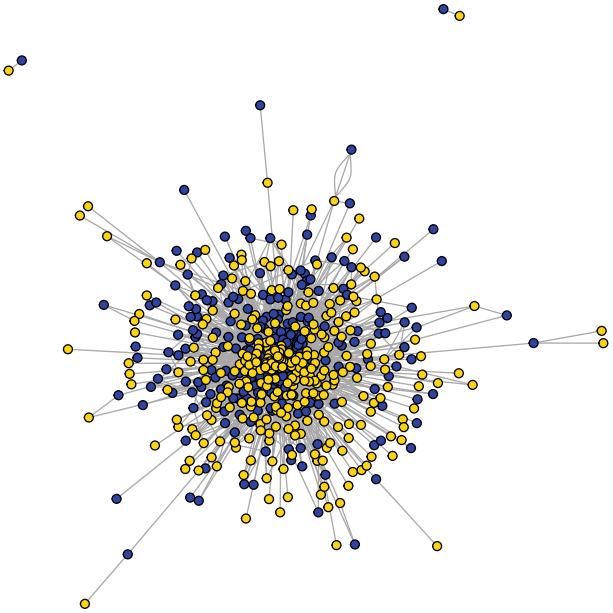


Fig. 4 Graph representation of data processed using the "split families" method, that considers separately husbands and wives in the marital network. The representation uses the default rendering algorithm in the igraph R package. "Husband" nodes are colored in blue and "Wife" nodes in gold

from a strictly pragmatic point of view, it is a simple way of treating a family's marriages to itself and marriages to other families equally. The side effect is that female and male nodes of the same family will be separated by, at least, one other node, unless there are intra-family marriages, of course; this is illustrated by node D in the figure, that has been separated in nodes $D-F$ and $D-M$, which are not directly connected. Please note that, in this method, *all* nodes and edges are changed, since "M" nodes can be connected only to "F" nodes.

To get some insight about the structure of the network with split families, the latter is rendered in Fig. 4 where male nodes are in blue and female nodes are in gold.[8] The figure shows how some families seem to occupy the center through the "husband" nodes, while others through their "wife" nodes, implying that some families achieve centrality by marrying their daughters (and providing dowry for it), while others, possibly more successful families, are sought for their position. Another interesting feature is that two small sub-networks, connecting a female of one family to the male of another, have been created; these sub-networks were originally connected through their "other" parts, so this is an artifact of this representation:

[8] As mentioned in the caption of Fig. 4, we are using the default node layout of the igraph package; as indicated in the manual, this method is called layout_nicely, and uses some heuristics to choose a specific layout based on the graph. Our point is that by incorporating new nodes and edges analogous to the existing ones, we do not need to add any special provision or layout algorithm to visualize the graph, except to highlight the newly added nodes.

Table 4 Top ten nodes in the Venetian dataset processed with the "split families" method according to the three centrality measures: betweenness (left), EV centrality (middle), and PageRank (right)

Split families					
Family	Betweenness	Family	EV centrality	Family	PageRank
Contarini-F	24660.309	Contarini-F	1.0000000	Contarini-F	0.0229260
Contarini-M	17793.624	Contarini-M	0.9448308	Contarini-M	0.0200763
Corner-M	12006.620	Morosini-F	0.6693481	Morosini-F	0.0140865
Querini-M	10241.281	Morosini-M	0.6545554	Morosini-M	0.0139162
Morosini-F	9048.820	Corner-F	0.5304401	Corner-M	0.0138219
Morosini-M	8715.207	Corner-M	0.4640765	Corner-F	0.0130296
Corner-F	8557.724	Querini-M	0.3866232	Querini-M	0.0121411
Querini-F	8225.212	Querini-F	0.3635297	Querini-F	0.0113944
Donato-F	7434.442	Priuli-F	0.3139506	Priuli-F	0.0099171
Donato-M	6456.792	Priuli-M	0.3014139	Donato-F	0.0096032

We cannot pretend that the female and male part of a family is disconnected even if there are not intra-marriages. However, trying to correct this effect would lead to additional artifacts.

Since differences in centrality values calculated on different networks are meaningless, to see how this method affects centrality measures we will compare family rankings. Top ten rankings for the split families network are shown in Table 4. Comparing them with the benchmark of Table 2 is not an easy task, due to the fact that the original nodes have been split and converted in others, but some facts emerge clearly nonetheless. The first, not surprising, is that the Contarinis, originally the most central family and also the one with the most intra-marriages, remain at the top of the ranking for both male and female. The second is that the family ranked fourth in the benchmark case, the Donatos, has now fallen (for both nodes) below the Morosinis who were originally only in fifth place. According to Table 3, the Donatos only had 11 inter-marriages, while the Morosinis had 23; this is why, the Donatos had to give way.

Other comparisons, however, are more difficult, because for the three centrality measure of Table 4, the male and female nodes, once split, end up in different position in the rankings, making it difficult to measure the centrality of the family as a whole. Moreover, from a historical or sociological point of view, it is very difficult to interpret the "male" and "female" members of the family as different actors of the social network. Therefore, although this methodology for taking into account intra-family marriages might open some interesting angles of research, we discard it.

Method 3: "Duplicated node"

The third method we propose is similar to the first one, because it creates a new node for each family with intra-family connections, which is why we will call it *duplicated nodes*. As with Method 1, this new node is connected to the original one by

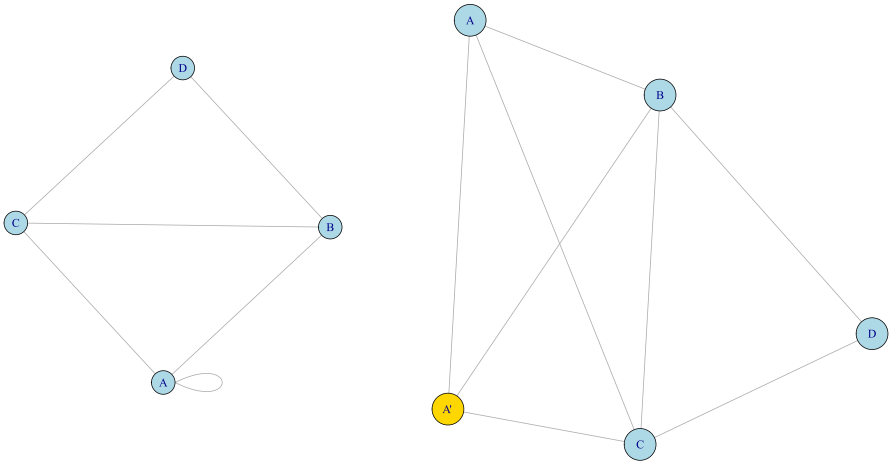


Fig. 5 Illustration of the "duplication" method, with the original network with a self-loop on the left and the modified network on the right. The original nodes are light blue and the "replica" ones gold

an edge of weight equivalent to the number of intra-family marriages; however, it is now also linked to all the nodes to which the original node was connected; thus, the "original" node and its "replica" have all the same connections (possibly including connections with "replica" nodes, of course), and are also connected to each other.

To better understand how this method works, Fig. 5 illustrates its effects over the usual simplified network with four nodes and one self-loop for node A: in practice, when A is duplicated to its "replica" A', if originally there was a registered marriage between, let us say, the Contarini (A) and the Morosini (B), an additional "fictional marriages" (that is, an edge) will be created between the other Contarini (A') and the Morosini; as well as, of course, the Contarini–Contarini (A–A') weighted edge. The justification is that, to account for intra-family marriages, we must consider that large families have two (undistinguished) parts; this would account for the links that have been created between the different parts of the two (large) families. Then, of course, intra-family marriages will link these two parts of the family, which again are undistinguished.

The Venetian network treated with the duplicated node method is shown in Fig. 6, with the "replica" nodes in gold; what we can observe in this image is that the nodes in gold, which are fewer in number that those in blue, are mainly located in the center of the graph. This tells us that families with intra-family marriages are more central than the others, i.e., they have higher centrality measures and are thus placed in the center by the layout algorithm.

We can now compare the centrality measures obtained by the duplicated node method with the benchmark. The new values are shown in Table 5, where the "replica" nodes have been eliminated from the ranking, because they have exactly the same values as the "original" one by design. A comparison of Tables 2 (left) and 5 (left) shows that the addition of the "replica" nodes decreased betweenness centrality for every node of the ranking. To understand why, consider that betweenness

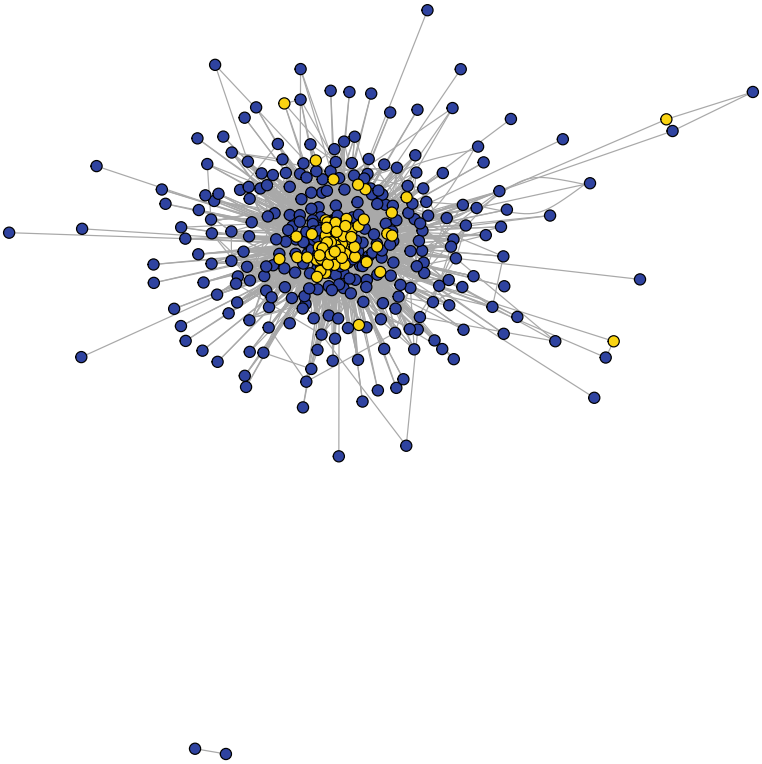


Fig. 6 Graph representation of data processed using the "duplicated nodes" method, that duplicates the node for families that have intra-connections; "original" nodes are colored in blue, "replicated" nodes in gold. The default rendering method in the igraph package has been used to place the nodes

Table 5 Top ten nodes in the Venetian dataset processed with the duplicated nodes method according to the three centrality measures: betweenness (left), EV centrality (middle), and PageRank (right)

Duplicated nodes					
Family	Betweenness	Family	EV centrality	Family	PageRank
Contarini	5354.666	Contarini	1.0000000	Contarini	0.0243540
Corner	3440.128	Morosini	0.7852031	Morosini	0.0165297
Querini	3117.479	Corner	0.6095175	Corner	0.0158097
Donato	2533.670	Querini	0.4723772	Querini	0.0135318
Morosini	1922.664	Priuli	0.3930271	Donato	0.0106255
Dolfin	1635.609	Dolfin	0.3513417	Priuli	0.0105549
Venier	1570.199	Giustinian	0.3497659	Malipiero	0.0096904
Michiel	1549.003	Michiel	0.3415891	Michiel	0.0093128
Balbi	1510.266	Zorzi	0.3361391	Loredan	0.0091636
Priuli	1496.089	Pisani	0.3302725	Zorzi	0.0089451

centrality measures how much a certain node is "in-between", that is, how often it is found when going from one random node to another using the shortest path. Our procedure has increased the number of nodes and edges, and thus, the measures for specific nodes are bound to be affected; in particular, what decreases per-node betweenness in the families shown in this ranking is the fact that other families have also duplicated their nodes, and this creates new nodes that will have the exact same short path passing through them; thus, the decrease in betweenness will be due mainly to the number of families with intra-marriages (duplicated nodes) that will still need to go through the node to get to other nodes.

However, a more proper way to compare two networks with different structures is to look at rankings and how they are affected by the new method. We see that the top six families in terms of betweenness centrality retain their position, but there are changes in the bottom four and, in particular, the Venier family, who did not belong to the original ranking, is now in seventh place. Overall, however, there are no drastic changes, which is good in this particular case, because this would have been at odds with the other family status indicators presented in the original paper.

The effect on eigenvector centrality observed when comparing Table 2 (center) and 5 (center) is also very small. The Contarini family, which we know has the highest number of intra-family marriages, is still the first in the ranking and the other families following the Contarini also maintain their position, with the exception of the tenth place which is now held by the Pisani instead of the Loredan. As we mentioned above for betweenness centrality, the small variations are consistent with the other status indicators for Venetian families. Notice also that the swap in places of the Pisani and the Loredan can be traced to the difference in internal marriages of the two families. From Table 3 we see that the Pisani family has eight self-marriages, while the Loredan only has four. Therefore, intra-marriages increase the (relative) centrality of the families who have them.

Finally, we turn our attention to PageRank centrality. This is a measure designed primarily for directed graphs; when used in undirected graphs, as we are doing here, it gives a recursive measure of influence alternative to eigenvector centrality. As can be seen by comparing the rightmost column of Table 2 and 5, for well-connected families, which are also those with a high number of intra-marriages, the changes are so negligible that the ranking of the first ten families remains the same.

Beyond changes in rankings, the effects of the duplicating nodes method on a family's influence can be illustrated using the simple network of Fig. 5. If the "replica" node (A') is interpreted as "another" part of the family that is not directly related (or not directly enough to prevent internal marriages) to "the original" one, when the new edges are added, an alternative way of getting to any part of the family, either the original node (A) or the other, separate, part of the family (A'), is created. The splitting family method, which divides the family by gender, was also a way of introducing "another" part of the family, but the division it introduced was fixed, whereas the duplicated node method simply indicates that there are different parts of the family, independent enough as to allow internal marriages, but still externally recognized as belonging to the same *casata* or dynastic house. Therefore, unlike the other methods studied so far, the duplicated nodes does have a straightforward interpretation in social terms.

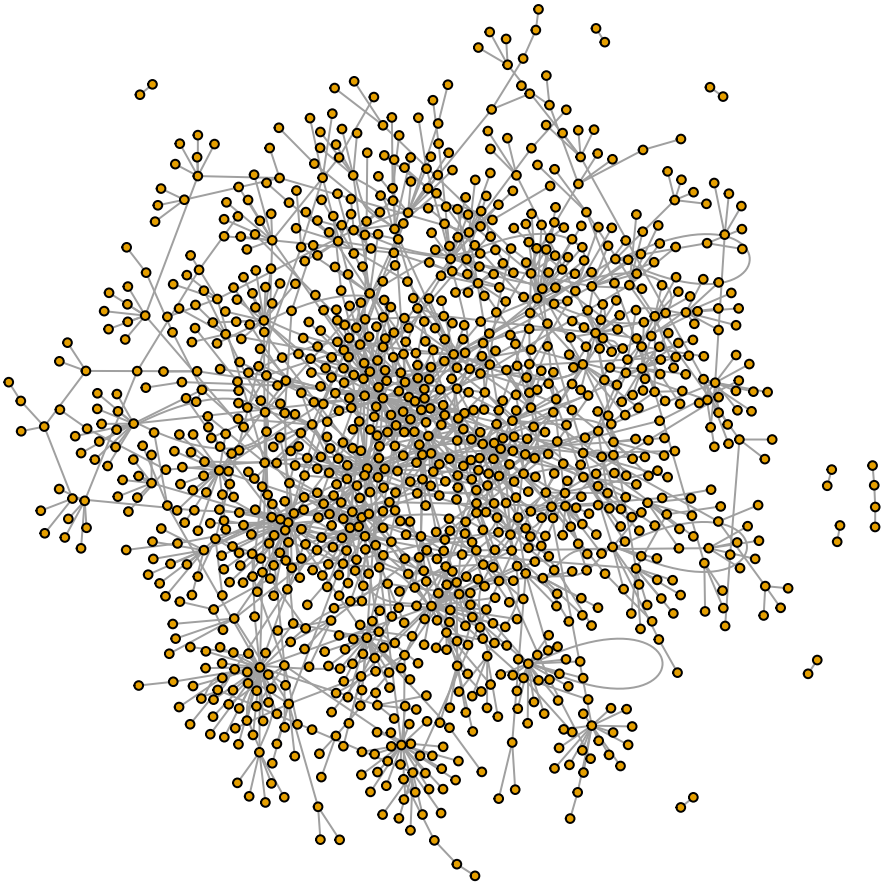


Fig. 7 The Taiwanese marriage network, using graphopt method for layout; this method is optimized for graphs with a large number of nodes

The duplicated node method and the Taiwanese network

We now apply the duplicated node method to the elite families marital network in Taiwan. The network is rendered in Fig. 7.

Table 6 lists all families with an intra-family marriage. The first two families in the table have two, the others only one. As indicated in "Datasets", this network has been extracted from Dluhošová [9] who classifies the most prominent families between "old" and "new" ones, as well as so-called "Mainland ruling elite" families. Additionally, using community analysis, the paper finds 11 communities, designated with letters from A to K, where letters are assigned in descending order of number of nodes; the "A" community, thus, is the largest with 4.33% of the nodes.[9] The main families in each community (referred to as "networks" in the paper) are also

[9] These communities cover 47% of the nodes, excluding networks that are not in the main component, for instance, nodes with degree = 1, and other ones that are simply too small.

Table 6 Intra-family marriages by family in the Taiwanese dataset

Family name	# Intra-family marriages
Qingshui Cai	2
Wu Huoshi	2
Chiang Kai -shek	1
He Yingqin	1
Liang Dingming	1
Lin Dingbang	1
Lin Weiyuan	1
Lin from Tainan	1
Liu Kuocai	1
Tainan Huang	1
Tainan Liu	1
Wang Yongqing	1
Xu Boyun	1
Xu Qiyou	1
Xu Shengfa	1
Yu Dengfa	1

Table 7 Top ten nodes in the Taiwanese elite families dataset with self-loops dropped according to the three centrality measures: betweenness (left), EV centrality (middle), and PageRank (right)

No self-loops					
Family	Betweenness	Family	EV centrality	Family	PageRank
Yan Yunnian	245000.97	Yan Fu	1.0000000	Chen Zhonghe	0.0130767
Lin Weirang	137273.67	Lin Weirang	0.7060325	Yan Fu	0.0126771
Lin Dingbang	137049.14	Wan Shao Mou	0.4005665	Yan Yunnian	0.0118801
Tainan Liu	108203.94	Yan Yunnian	0.3280270	Cheng Zhong Mo	0.0082196
Gu Xianrong	94485.35	Jia Dehuai	0.2505374	Tainan Liu	0.0081931
Chen Zhonghe	88091.43	Liang Jinying	0.2467246	Wu Huoshi	0.0081443
Lin Bo	87328.00	Chen Zhonghe	0.2332053	Wu Xiuqi	0.0078990
He He Yi	86993.12	Du Chongming	0.1917979	Lin Weirang	0.0077906
Chen Shoushan	84040.52	Gu Xianrong	0.1880376	Taoyuan Wu	0.0076716
Lin Xiantang	73326.43	Chen Baochen	0.1741956	Cathay Cai	0.0075025

indicated by a number in descending order of centrality: for instance, the Ling Xiantang family, labeled A1, is the most prominent family in the "A" community, the "Wufeng Ling family network".[10]

[10] Sometimes the Lin Xiantang family is also referred to as Wufeng Ling "upper" branch in the paper. We would like to point out that the way families are identified is not uniform across the two social net-

As we did before, we first compute benchmark centrality measures when self-loops are eliminated. The resulting rankings for the ten top families are shown in Table 7 for betweenness centrality, eigenvector centrality, and and PageRank, respectively. Looking at the three tables, we immediately see a striking difference between the Taiwanese network and the Venetian one: in the benchmark measures for Venice (Table 2), the most central family is the Contarini, the same for all three rankings. Furthermore, seven out of ten families appear in all three rankings and a total of only 13 families are found across the three rankings, i.e., all measures of centrality identify the same small group of families. In the Taiwanese dataset, instead, the first spot in the ranking is taken by a different family for each measure, and a family ranked first for one measure is not even in the top ten for another, as in the case of the Yan Fu family, which is first in eigenvector centrality but not in the top ten for betweenness centrality. Looking at the data, one notices that the structure and composition of the network are totally different, and the network is much more sparse, which causes different measures to increase the centrality of specific families depending not so much on their degree, but on the position they have in the network; this is also noticeable in the betweenness centrality measures, which in this case are two orders of magnitude greater than in the Venetian network.

Having established in the previous section that the duplicated node method can provide us with a way to introduce self-loops in any of these centrality measures, let us then apply it to this network. As we can see from the rendering of the resulting network, shown in Fig. 8, duplicating nodes only adds 16 new nodes in a network with numerous families; this is in sharp contrast with the Venetian network of Fig. 6 which, as we said, had less nodes and proportionally many more self-loops. Therefore, this is obviously an extreme case of network with intra-family ties: a very small percentage of families are big enough to allow internal marriages. As the following analysis shows, they do, however, have an impact on centrality measures.

Let us see how the introduction of these new ties impacts the rankings of the three centrality measures considered. The results are shown in Table 8. A first look indicates that even though we have only introduced 16 new nodes with their edges, there is a considerable effect in all the rankings. However, the impact varies from one ranking to the next. We compare the benchmark and the new ranking in turn.

Starting with betweenness centrality and comparing Tables 8 (left) and 7 (left), we see that only the top two families and the sixth keep their position in the ranking. In addition, one family, Tainan Liu, which in the benchmark case was included in the top ten in fourth place, drops out of the new ranking. The fact that there are small corrections in the families included in the top ten ranking matches what happens with the Venetian marital network. Another common feature is the decrease in betweenness centrality values, although the change is now

Footnote 10 (continued)

works we used in this paper. For example, in the Taiwanese network Wufeng Ling "upper" branch (A1, Lin Xiantang) and "lower" branch (A2, Ling Dingbang) are regarded as different families, and thus correspond to two different nodes. In the Venetian Republic network, instead, all branches of a family are assigned to a single node. This could be due to cultural differences or simply lack of data, and it possibly explains the difference in the number of self-loops in the two datasets.

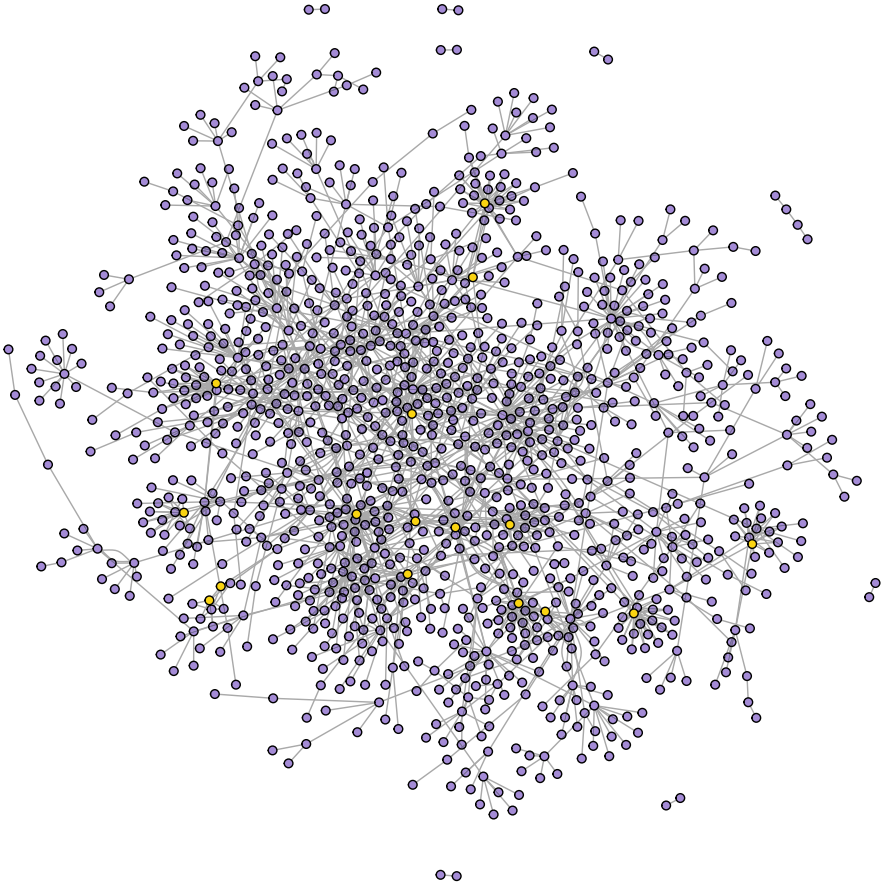


Fig. 8 The Taiwanese network using the "duplication" method; "original" nodes are colored in blue and "replicated" nodes in gold. The graphopt method has been used to place the nodes, as in the previous figure

on a different scale: in the Venetian network, the value was almost halved, while in this network, there is only a small correction; this can be explained by the small number of families with self-loops, which implies that the information we are adding to the network should not have such a big *numeric* impact, although, as we have seen, it has an impact in the rankings.

What really sets this case apart is that the changes are not induced *directly* by the introduction of new nodes and edges, because the families that change their position in the ranking are not those with self-links; in fact, there are only two families with (a single) intra-marriage, among those that are in the top ten for betweenness centrality; these are Lin Dingbang and Tainan Liu, and for both betweenness centrality *decreases*, for the Tainan Liu family to the point that it drops out from the ranking.

Table 8 Top ten nodes in the Taiwanese dataset processed with the duplicated nodes method according to the three centrality measures: betweenness (left), EV centrality (middle), and PageRank (right)

Duplicated nodes					
Family	Betweenness	Family	EV centrality	Family	PageRank
Yan Yunnian	244945.45	Tainan Liu	1.0000000	Chen Zhonghe	0.0127051
Lin Weirang	135176.41	Qingshui Cai	0.8073431	Yan Fu	0.0124040
Gu Xianrong	95242.51	Lin Xiantang	0.5289344	Yan Yunnian	0.0112790
Lin Bo	92276.96	Lin Dingbang	0.3718102	Cheng Zhong Mo	0.0077594
He He Yi	90802.57	Lin Weiyuan	0.3359714	Wu Xiuqi	0.0075584
Chen Zhonghe	87476.25	Cheng Zhong Mo	0.2761809	Lin Weirang	0.0074943
Chen Shoushan	84022.44	Xinzhu Zheng	0.2511753	Taoyuan Wu	0.0072899
Lin Dingbang	74059.65	Huang Chaoqin	0.2424090	Cathay Cai	0.0070953
Lin Xiantang	73834.25	Li Chunsheng	0.2321755	Chen Lingan	0.0062887
Shi Fengxiang	69282.68	Zhu Ying San	0.2307947	Tainan Liu	0.0061754

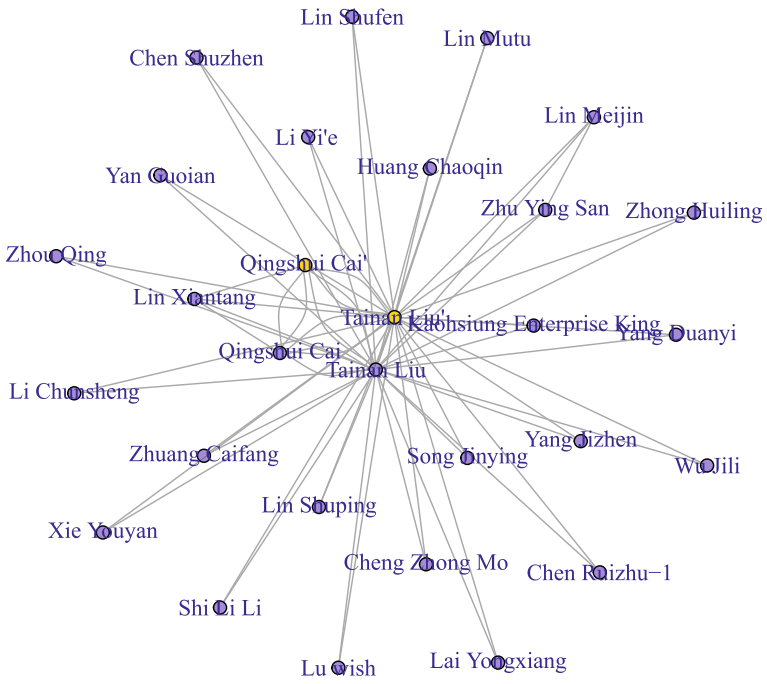


Fig. 9 *Ego* network for the Tainan Liu family

Since the Tainan Liu family is also the one that tops one of the other two rankings, we should probably analyze its position in the network to illustrate how it achieves it and how it is impacted by the addition of new nodes; the ego network, that is the sub-network that includes all nodes connected to Tainan Liu, is shown in

Fig. 9. As it can be seen, it is (mostly) a star-type network: The Tainan Liu family serves as connection for a good number of nodes, and most of them can only connect to the rest of the network through it. The ego network also includes 2 of the 16 nodes that have been added; in such a sparsely connected network, the addition of new nodes and their corresponding edges is bound to have a great local impact. The fact that many nodes are only connected through the Tainan Liu family explains its high eigenvector and Page Rank centrality; on the other hand, since another node has been added in this version of the network, it provides alternative paths to the many nodes connected to it, thus decreasing its centrality, critically in this case, since it makes the family drop from the eigenvector centrality ranking. This is a totally intended effect of the introduction of new nodes and edges for families with intra-marriages.

If we compare Tables 7 (center) and 8 (center), which show the rankings in terms of eigenvector centrality, the situation is different: The Tainan Liu family, which was not among the top ten families in the benchmark case, rises to first place in the duplicated nodes ranking; moreover, all the top ten families in the new ranking were not in the benchmark top ten. Also notice that four out of the top five families in the new ranging have self-loops (all of them except Lin Xiantang).

Since the ranking has totally changed, we need to ground these results to what was published in the original paper. This is not straightforward, because, as explained at the beginning of this subsection, families were classified along two dimensions, size and centrality, represented as letters and numbers, respectively. In what follow we will try to use this classification to understand what type of families are raised to the top in the duplicated node ranking. Looking at the top five families for eigenvector centrality in the new ranking (center column of Table 8), we find, in descending order, Tainan Liu family (A3), Qingshui Cai (A7), Lin Xiantang (A1), Ling Dinbang (A2), and Ling Weiyuan (F1). On the other hand, in the benchmark ranking without self-loops, in the top two positions there are, in descending order, Yan Fu (F3) and Lin Weirang (F2), followed by Wan Shao Mou, which is not even listed in the largest or most prominent families,[11]) and then Yan Yunnian (B1) and another unlisted family, Jia Dehuai.

Since the most prominent families by size are identified with the first letters of the alphabet and by centrality with smaller numbers, the duplicated node method seems better able to identify the most prominent families. In other words, looking at eigenvector centrality, the ranking without self-loops put seemingly more irrelevant families, either from smaller communities or less central) at the top, while the new ranking picks families from the A cluster and with small numbers.

Finally, the PR rankings, shown in Tables 7 (right) and 8 (right), do not change too much when including self-loops in the way of duplicated nodes (as it was the case with the Venetian marriage network). The top four families in the ranking do not vary: Chen Zhonghe is J1 (unnamed in the original paper), Yan Fu (F3), Yan Yunnian (B1), and Cheng Zhong Mo (A12) families, scattered over different networks. Wu Xiuji (B10) has substitute of Tainan Liu (A3) in fifth place, and Lin Weirang (F1) in the next one. The B group corresponds to the Jilong Yan network,

[11] See p. 139 of [9].

Table 9 Summary of features of the three methods we have introduced for accounting for intra-family marriages

Method	Structural changes	Community changes	Artifacts
New nodes	No	No	Yes
Split families	Yes	Yes	Yes
Duplicated nodes	No	(Yes)	No

also composed by *old* families. This change in ranking is certainly more difficult to interpret, since both A and B families are *old*. However, it is interesting to note that the Lin Weirang family, which is mentioned in Dluhošová's paper, section 5.1, as the link of the F network (Banqiao Lin, old) with the C network (ROC Mainland ruling family, mainland elite families) and the J network (Gaoxiang Chen, old), has been boosted with respect to the benchmark ranking, where it was in eighth place. Although not a family including self-loops itself, this does show how the structural changes induced by the duplicated method can make the resulting PageRank better reflect the actual status of a single node within the network.

Discussion

In this section, we discuss the soundness of the three methods and how they affect various dimensions of the graph on which they are applied. In particular, we will consider four aspects: first, whether the method introduces structural changes at the node level; second, whether and how the new edges brings changes at the community level; third, whether it produces artifacts that have no plausible explanation; and, fourth, how it accounts for the importance of intra-family marriages. Table 9 contains a summary of our conclusions, based on the results of "Methods and results".

Looking at the checklist shown in Table 9, the "duplicated nodes" method of "Method 3: "Duplicated node"" stands out above the other two. First, it does not introduce structural changes at the node level, since the "replica" nodes are structurally equivalent to the original ones, whereas the new nodes in the "split family method" are not. Second, it does introduce changes in the community, due to the new edges, but this effect is intentional and obviously introduces global structural changes, as we have seen in the previous section. This result is also achieved by the "split family method", but not by the "new nodes" method which, by adding a single weighted edge for each family with intra-marriages, fails in this respect. Finally, what really distinguishes the "duplicated nodes" method from the other two is that it does not create unexplained artifacts: the new nodes it introduces can be interpreted as *other* representatives of the family. In practice, the duplicated nodes method takes into account the fact that families with some amount of intra-family marriages are, by custom or law, large enough to have more than one "actor" or to have agency in several directions. However, the two nodes of the same family will be, from the analytic point of view, indistinguishable from each other, and externally considered

the same, which is why they have exactly the same values for all centrality measures. The fact that they are linked to each other also accounts for their intra-family marriages, and explains how the latter contribute to family cohesion.

The *new nodes* method (the "Method 1: "New nodes""), by contrast, introduces other nodes that are only linked to the original ones, and thus have no real interpretation, and the "split nodes" method (the "Method 2: "Split families""), by dividing every family along gender lines, treats the male and female parts of the family as separated agencies and introduces the undesirable artifact that there might be two nodes of the same family that are not linked to each other, as only families with self-loops have such a link. Moreover, with the "split family" method, measures of centrality are structurally different for female and male nodes of the same family, and thus impossible to ground in the family social reality, as explained above.

All three methods proposed are heuristic, and their validity can only be ascertained post hoc, by empirically verifying whether they better represent the social or historical status and position of the families. Of course, self-links could be also investigated in other ways outside the field of social network analysis; for instance, we could simply look at the percentage of intra-family marriages versus inter-family marriages. However, if we acknowledge that marriages form social links, social network analysis offers the researcher quantitative insights at the actor (family) and meso (community) level that would otherwise not be available, and excluding intra-family marriages from it could lead to quantitative errors that are difficult to overcome.

Conclusion

Intra-family ties are an important part of the dynamics of marital networks; however, they have so far rarely been taken into account when computing centrality measures. In this paper, we propose alternative ways to incorporate these intra-family ties into the graph representing the marriage network; all of the methods suggested are compatible with the use of existing social network analysis software.

After empirically testing three alternative methods with two (very different) marriage datasets, one from the Republic of Venice and the other from Taiwanese elite families across the 19th–20th centuries, we conclude that the most meaningful way to introduce self-loops is the "duplicated nodes" method. This creates a "replica" node for each family with non-zero internal marriages and connects it to the original node with an edge whose weight is equal to the number of internal marriages in the family; it also connects the "replicas" node to all other nodes to which the original node was connected. The advantage of this method is that it makes it possible to estimate the influence of intra-family marriages into the structure of the whole network, without producing artifacts. These "replica" nodes share the same actor-level measures with the node they mirror and have a well-founded meaning, as they can be interpreted as "another representation" of the family.

We show that the "duplicated nodes" method is able to create family rankings, with respect to the three main centrality measures we consider that better

reflect the family social status, as reported by the original paper in which the two networks were first introduced. This is also the case with a very small percentage of self-loops, whether they are calculated with respect to the total number of marriages or the total number of families.

Another advantage of the method is that it is not computationally intensive and can be performed either manually, by manipulating the spreadsheet, or using any data-oriented scripting language such as R or Python. In the near future, we plan to publish a library, using the language R and possibly other languages, to allow this method to be easily used, so that as more marital network datasets become available, it will be possible to study their general dynamics even in the presence of extensive intra-family ties.

An additional improvement to this method would be to adjust the weights of the new edges, so that we are able to approximate, in the case of EV centrality, its value computed using self-loops. That way we could validate numerically, as well as qualitatively, the results obtained. Additional validation can be reached by applying the duplicated nodes method to other networks with self-loops, such as the ones mentioned in the "Introdcution" and "A survey of marriage networks".

Author Contributions JJM and MCM have written the paper and performed the analysis shown in it, JJM wrote the code embedded in the paper. MCM has performed extensive revision and suggestions, and proposed new analysis and hypotheses.

Funding Funding for open access publishing: Universidad de Granada/CBUA. This work is supported by the Ministerio Español de Economía y Competitividad (Spanish Ministry of Competitivity and Economy) under Project PID2020-115570GB-C22 (DemocratAI::UGR). Funding for open access charge has been provided by the Universidad de Granada/CBUA.

Availability of data and materials Data are available from https://github.com/JJ/Intra-family-networks/ in R data format, and originally from https://diegopuga.org/data/venice/ and https://doi.org/10.6084/m9.figshare.12639104.v1. License is the same as the rest of the project, that is, GPL.

Code Availability Code is embedded in the source of the paper at https://github.com/JJ/Intra-family-networks/blob/main/paper/main.Rtex. (Note to reviewers: so as not to interfere with the review process, this repository is for the time being private.) As the rest of the paper and project, it is available under the GPL license.

Declarations

Conflict of interest There is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Luke, N., Munshi, K., & Rosenzweig, M. (2004). Marriage, networks, and jobs in third world cities. *Journal of the European Economic Association, 2*(2–3), 437–446. https://doi.org/10.1162/154247604323068122
2. Kuper, A. (2001). Fraternity and endogamy. The House of Rothschild. *Social Anthropology, 9*(3), 273–287. https://doi.org/10.1111/j.1469-8676.2001.tb00153.x
3. Landherr, A., Friedl, B., & Heidemann, J. (2010). A critical review of centrality measures in social networks. *Business and Information Systems Engineering, 2*(6), 371–385. https://doi.org/10.1007/s12599-010-0127-3
4. Merelo, J. J., & Molinari, M. C. (2023). Self-loops in social networks: behavior of eigenvector centrality. In *Proceedings WIVACE 2023, to be published*
5. Chartrand, G., & Zhang, P. (2013). *A first course in graph theory*. Courier Corporation.
6. Boesch, F., & McHugh, J. (1974). Synthesis of biconnected graphs. *IEEE Transactions on Circuits and Systems, 21*(3), 330–334.
7. Shafie, T. (2016). Analyzing local and global properties of multigraphs. *The Journal of Mathematical Sociology, 40*(4), 239–264. https://doi.org/10.1080/0022250X.2016.1219732
8. Shafie, T. (2015). A multigraph approach to social network analysis. *Journal of Social Structure, 16*(1), 1–21. https://doi.org/10.21307/joss-2019-011
9. Dluhošová, T. (2020). Marital networks and portfolios of prestige: Digital humanities perspectives on the study of Taiwanese elites. *European Journal of East Asian Studies, 19*(1), 124–160. https://doi.org/10.1163/15700615-01901003
10. Catino, M., Rocchi, S., & Vittucci Marzetti, G. (2022). The network of interfamily marriages in 'Ndrangheta'. *Social Networks, 68*, 318–329. https://doi.org/10.1016/j.socnet.2021.08.012
11. Csardi, G., & Nepusz, T. (2006). The `igraph` software package for complex network research. *InterJournal, Complex Systems, 1695*(5), 1–9.
12. Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. In *Proceedings of the international AAAI conference on web and social media* (vol. 3, pp. 361–362). https://doi.org/10.1609/icwsm.v3i1.13937
13. He, M., Glasser, J., Pritchard, N., Bhamidi, S., & Kaza, N. (2020). Demarcating geographic regions using community detection in commuting networks with significant self-loops. *PLoS ONE, 15*(4), 1–31. https://doi.org/10.1371/journal.pone.0230941
14. Zamborain-Mason, J., Russ, G. R., Abesamis, R. A., Bucol, A. A., & Connolly, S. R. (2017). Network theory and metapopulation persistence: Incorporating node self-connections. *Ecology Letters, 20*(7), 815–831. https://doi.org/10.1111/ele.12784
15. Saura, S. (2018). Node self-connections in network metrics. *Ecology Letters, 21*(2), 319–320. https://doi.org/10.1111/ele.12885
16. Smith, M. B., Blakemore, J. K., Ho, J. R., & Grifo, J. A. (2021). Making it (net)work: A social network analysis of "fertility'' in twitter before and during the COVID-19 pandemic. *F &S Reports, 2*(4), 472–478. https://doi.org/10.1016/j.xfre.2021.08.005
17. Lichoti, J. K., Davies, J., Kitala, P. M., Githigia, S. M., Okoth, E., Maru, Y., Bukachi, S. A., & Bishop, R. P. (2016). Social network analysis provides insights into African swine fever epidemiology. *Preventive Veterinary Medicine, 126*, 1–10. https://doi.org/10.1016/j.prevetmed.2016.01.019
18. Merelo-Guervós, J. J., & García-Valdez, M. (2022). Agile (data) science: A (draft) manifesto. https://doi.org/10.48550/arXiv.2104.12545
19. Padgett, J.F., & Ansell, C. K. (1993). Robust action and the rise of the Medici, 1400–1434. *American Journal of Sociology 98(6),* 1259–1319 . https://www.jstor.org/stable/2781822
20. Puga, D., & Trefler, D. (2014). International Trade and Institutional Change: Medieval Venice's Response to Globalization. *The Quarterly Journal of Economics, 129*(2), 753–821. https://doi.org/10.1093/qje/qju006
21. Telek, Á. (2017). Marrying the right one—Evidence on social network effects in politics from the Venetian Republic. https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=SAEe2017 &paper_id=520
22. Merelo-Guervós, J. J. (2022). What is a good doge? Analyzing the patrician social network of the Republic of Venice. arXiv. https://doi.org/10.48550/ARXIV.2209.07334. arXiv: 2209.07334

23. Batagelj, V. (1996). Ragusan families marriage networks. In: A. Ferligoj, A. Kramberger (Eds.) *Develop. in Stat. and Methodology*. Metodoloki zvezki, vol. 12. Ljubljana. http://dk.fdv.uni-lj.si/MetodoloskiZvezki/Pdfs/Mz12Batagelj.pdf
24. Lee, S., & Lee, W. (2017). Strategizing marriage: A genealogical analysis of Korean marriage networks. *Journal of Interdisciplinary History, 48*(1), 1–19. https://doi.org/10.1162/JINH_a_01086
25. Tackett, N. (2020). The evolution of the Tang political elite and its marriage network. *Journal of Chinese History, 4*(2), 277–304. https://doi.org/10.1017/jch.2020.6
26. Cruz, C., Labonne, J., & Querubin, P. (2017). Politician family networks and electoral outcomes: Evidence from the Philippines. *American Economic Review, 107*(10), 3006–3037. https://doi.org/10.1257/aer.20150343
27. Haim, D., Nanes, M., & Davidson, M. W. (2021). Family matters: The double-edged sword of police-community connections. *The Journal of Politics, 83*(4), 1529–1544. https://doi.org/10.1086/715071
28. Naidu, S., Robinson, J. A., & Young, L. E. (2021). Social origins of dictatorships: Elite networks and political transitions in Haiti. *American Political Science Review, 115*(3), 900–916. https://doi.org/10.1017/S0003055421000289
29. Battaglini, M., & Patacchini, E. (2019). Social networks in policy making. *Annual Review of Economics, 11*, 473–494. https://doi.org/10.1146/annurev-economics-080218-030419
30. Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry, 40*(1), 35–41. https://doi.org/10.2307/3033543
31. Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks, 29*(4), 555–564. https://doi.org/10.1016/j.socnet.2007.04.002
32. Ruhnau, B. (2000). Eigenvector-centrality-a node-centrality? *Social Networks, 22*(4), 357–365. https://doi.org/10.1016/S0378-8733(00)00031-9
33. Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems, 30*(1), 107–117. https://doi.org/10.1016/S0169-7552(98)00110-X

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.