ColNes PUBLISHING

SHORT COMMUNICATION

Kinship network evolution in Argentina. An exploration based on online data

Juan M. C. Larrosa[I, II,]*, **Franco Galgano**[III], **Emiliano Gutiérrez**[I, II]

[I] Departamento de Economía, Universidad Nacional del Sur, Argentina.
[II] Instituto de Investigaciones Económicas y Sociales del Sur (IIESS), Argentina.
[III] Departamento de Física, Universidad Nacional del Sur, Argentina.

* Corresponding author:
    Email: jlarrosa@uns.edu.ar

ABSTRACT

We present the initial stages of a research project for a big data genealogical study. Consanguineous ties among members of a family form kinship networks. Family is the first social network for actors and historically was the source of education, support, and socialization. Family is still the source of power in the royalty, and family business has commonly been the first stage for future big firms. Nepotism is still a usual practice in politics and civil service. We analyze contributions to population evolution and propose to empirical test theoretical results. For that to be accomplished, we gather online data and present an initial descriptive analysis. In the end, following previous contributions, we also extend data usage for identifying historical lineages from colonial times in Argentina and analyzing if that antecedent portrays different benefits to descendants.

Keywords: kinship network; genealogical data; family type; online data; Argentina.

1. INTRODUCTION

Families remain the first social network anyone comes across in life. The mother and the father are the very early social contacts a human still makes the following, if the case allows, by brothers or sisters and the available bloodlines. Although the role of familial ties in post-modern human interactions may appear somehow devaluated —especially from a Western-world perspective— consanguinity and marriage relations (Bailey *et al.,* 2014; Walker and Bailey, 2014) still determine, for a large part of Mankind, the social core of interactions (Alger and Weibull, 2010; Schulz, 2022). Kinship relations are crucial to the founding theories of structural anthropology, as they grant the basis of social cohesion, alliance, and reciprocity (Roth *et al.,* 2013; Shenk *et al.,* 2016).

However, when considering the long-term evolution of families, how do different properties play a role in the sustainability of populations? We review two main theoretical contributions that base their conclusions on numerical simulations. They present an initial population assuming default values for critical parameters and let the simulation run and derive final values. The results are shown in the following section. A later debate emerges on whether an empirical work could confirm the simulations. For further analyzing the question, we organize the gathering of genealogical data. As the simulations assume large populations over long periods, we require data accordingly. Then, we rely on massive online databases. Online data can be accessed by dedicated algorithms that systematically copy, store, and

organize a large amount of data. The following contribution shares the experience of the process for a genealogical database, explore initial questions and presents future paths of research given the data availability.

2. MOTIVATION AND GOALS OF THE PROJECT

A formal approach to genealogical or kinship networks is provided by Zanette (2019). Martinez Alcalá and Zanette (2022) extend the previous contribution by assuming growing and decaying populations. Both contributions model kinship networks between heterosexual monogamous men and women. Relationships are established between consanguineous or political brothers and sisters. A random parameter determines the number of consanguineous siblings. The only condition for marriages is the exclusion of brother-sister ties. Numerical simulation studies: number and size of connected components, distribution of degrees, assortativity, distances within the giant component. It is compared in each case with a random network with the same degree distribution. It is stated that these social networks have typical behaviors of random networks. However, the second contribution studied the structural properties of a class of model social networks that represent relations between consanguineous brothers and brothers-in-law in the case where the size of the married population varies between successive generations. These networks are characterized by Poissonian degree distributions and a connected component that brings together most of the nodes. It also has high clustering and assortative values. Numerical simulations were performed for their study and compared with Erdős-Rényi networks of the same size and connectivity. Global clustering and variety were shown to be high unless the size of the married population declined dramatically. By contrast, the most significant connected component collapses when the married population shrinks to almost two-thirds of its size in the previous generation. We also consider the early contribution of Derrida *et al.* (1999).

We pose the question: Do empirical data corroborate these regularities? We propose three actions that we only describe more thoroughly in this contribution. Firstly, to obtain empirical evidence to test the hypothesis 1) Does the empirical evidence verify the simulated results? And if the evidence demonstrates, a second hypothesis emerges: Are these validations context-specific?

The other two courses of action are i) to replicate the Zanette (2019) and Martínez Alcalá and Zanette (2022) models in the context of agent-based simulation models, and finally, to mathematically model both models to extend and formalize the detected stylized facts.

3. DATA AND ANALYSIS

Our project uses web scraping techniques to gather massive data from websites. The algorithm points to a fixed position on a page and copies and stores the information. By visiting diverse web pages, it can obtain a complete record. We collected data from 163 thousand families, including 225 thousand women and 232 thousand of men. Data range from 1540 to 2020.

Data were extracted from an online database by web scraping.[1] Specific data manually removed from the site is presented in a previous contribution (del Valle y Larrosa, 2019): An example of a family network is shown in Figure 1. Kohler and Hammel (2001) and Murphy (2011) are other contributions in the same spirit. The complete database lacks completeness in many fields: absence of a date of birth (majority) and date of death (fewer cases). However, we detected a strong presence of wedding dates that made it more plausible to use families as the unit of analysis.

Nonetheless, most of the data is undated. Only 36% are geographically or temporally identified or both. Nevertheless, by experience from previous works, we developed different imputation strategies for "herding" unidentified data to the flock. For instance, when an individual is related to another geographically and temporally identified, it is thinkable that they might have shared both: In that case, imputation might apply. The process of data cleansing has been highly demanding in time and resources.
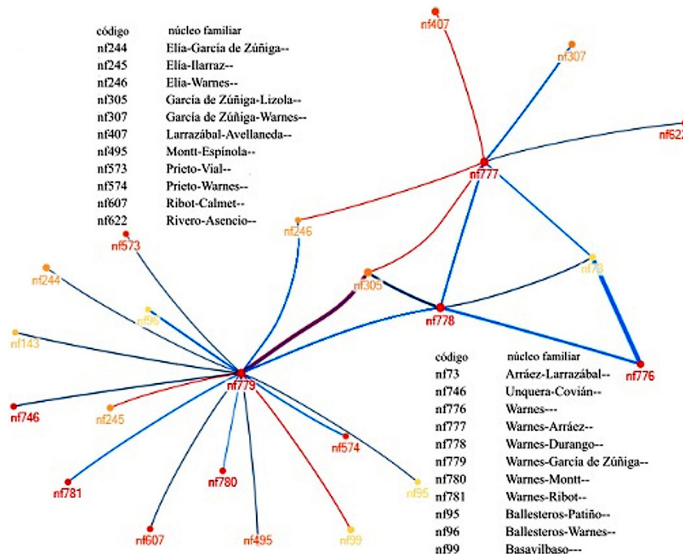
---

[1] http://www.genealogiafamiliar.com.ar

AWARI
Vol. 3, 1-8, DOI: 10.47909/awari.150

Kinship network evolution in Argentina.
An exploration based on online data

Figure 1. Example of one family from Del Valle and Larrosa (2019).

It is worth noting that the database is not an official record. The owners of the site add records, but also any user can also contribute. A potential selection bias is present because the record upload is not a random nor a stratified process. The number of weakly connected components is 140,686. It is understandable, given that links are rigidly formed by blood or marriage. The largest, most significant component is 464 vertices (0.104%). The number of strongly connected components reaches 292,079, while the most significant component, in this case, is 198 vertices (0.44%). Watts-Strogatz clustering coefficient is 0.6874, and the network clustering coefficient (transitivity) reaches 0.3208. An approach for cluster identification, such as a simple variant of the Louvain algorithm, grants 140,686

clusters with an estimated modularity of 0.999. The density is 0.000003, and the average degree is 2.72893387 (1.364), again understandable given the context. The diameter is 40, and the average path length is 1.887. In the case of assortativity measures rank, we can observe: input-input degree assortativity=0.04398; output degree assortativity=0.02959, output-input degree assortativity=0.26247; and output-output degree assortativity=-0.00862. It is slightly assortative in output-input: nodes with higher outdegree tend to be associated with nodes with high indegree. The network betweenness centralization is 0.0000003. Family degree distribution seems to follow a free-scale architecture at a glance (Figure 2), but that has not been confirmed with log-log transformation and power-law fitting.
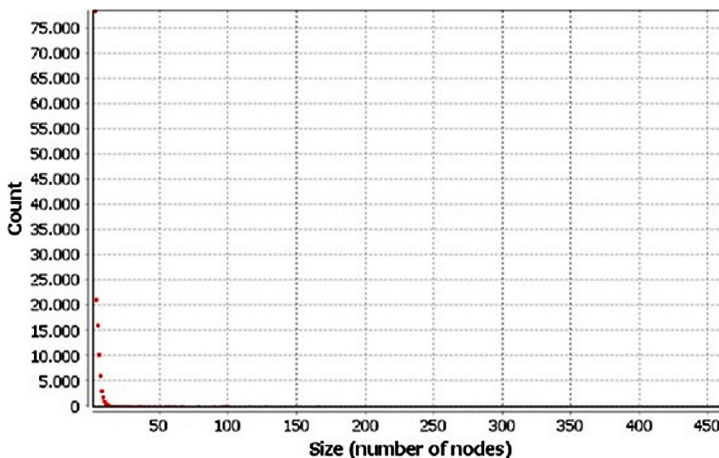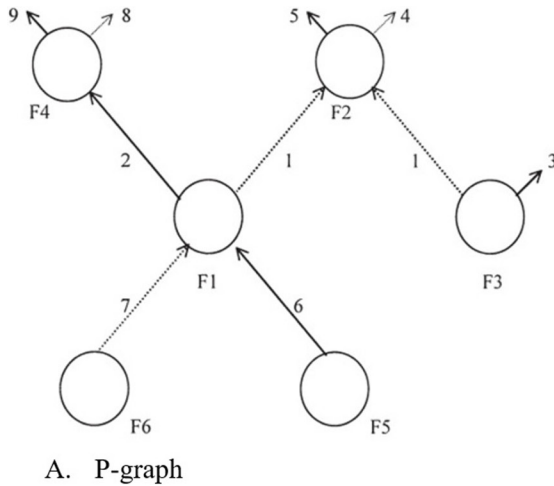

Figure 2. Degree distribution.

3.1 Type of family graph

Generic literature on genealogy states how a unit of analysis can be modeled (White and Jorion, 1996; White *et al.,* 1999). Two primary forms of modeling family are followed: P-graphs and Ore-graphs. P-graphs assume families as nodes and edges as siblings from one family to the other (Figure 3.A). Ore-graphs make any node to be created by two precedents, and no node is allowed to create an ancestor (Figure 3.B).
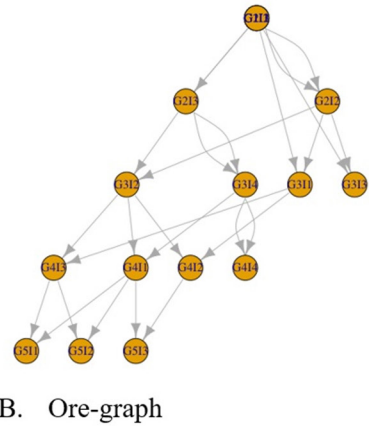
A.   P-graph

B.   Ore-graph

Figure 3. P-graph and Ore-graph.

In our case, we do not choose the alternative of modeling families by considering matrimony as one node. We model spouses as separate entities, maintaining their individuality. However, to keep the graph acyclic, we define matrimony as a direct link between the husband and wife. Both then directly to their siblings (Figure 4). This is a rather Patriarchal approach but is followed for a technical reason.
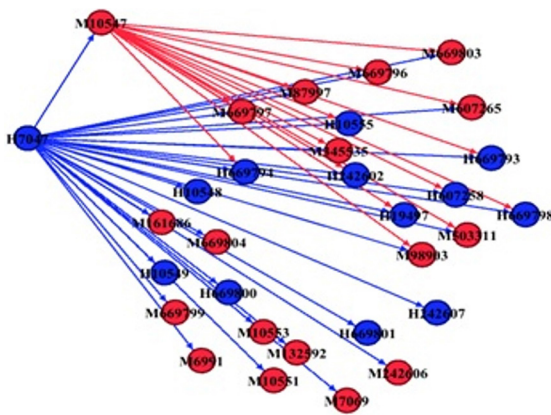
Figure 4. An example of a family.

By creating family graphs following this approach, we can grant that graph is acyclic except if endogamy emerges.

We begin by sorting families by region. We focus mainly on Argentina, where the database also has data from other countries. Table 1 presents data obtained from the main Argentinian cities. Given its importance, Buenos Aires represent a main logical historical data provider. However, the distribution of families across cities in the period diverges from the current demographic composition compared to the population share obtained in INDEC (2012). There is a natural explanation given that they are cumulative data, and many cities were not even founded when other cities were well-organized urban centers by that time.

3.2 Time and geography in the sample

Table 2 presents the composition of men, women, and family sizes in Buenos Aires, Córdoba, and Mendoza. In the sample, Cordoba and Mendoza show a smaller average family size than Buenos Aires. Gender composition is mildly biased towards male predominance. In any case, the average family size for Buenos Aires rounds 5 members, for Córdoba rounds 4 members, and for Mendoza is one step lower by rounding 3 members.

AWARI
Vol. 3, 1-8, DOI: 10.47909/awari.150

Kinship network evolution in Argentina.
An exploration based on online data

City/Province	Families	Sample Percentage	Actual Percentage*
Buenos Aires	13.101	56,1%	24,8%
Córdoba	3.258	14,0%	11,4%
Rosario	2.358	10,1%	8,1%
Mendoza	1.689	7,2%	14,9%
Tucumán	1.065	4,6%	6,8%
Salta	853	3,7%	4,6%
Luján	313	1,3%	0,9%
La Plata	190	0,8%	6,8%
Santiago del Estero	175	0,7%	2,4%
Jujuy	147	0,6%	5,8%
La Rioja	86	0,4%	2,9%
Mar del Plata	50	0,2%	5,0%
Bahía Blanca	47	0,2%	2,8%
Posadas	19	0,1%	2,7%

Source: The Authors.
* Percentage obtained from the 2010 Census data (INDEC, 2011).

Table 1. Geography of identified families.

Data among families are diverse. Many actors are well-documented, but the majority have less than well-documented attributes. For instance, Table 3 presents three specific cases. Red lines refer to the ego's direct links. All ego networks show 4.5 depth inside the network: 1 depth means direct links, 1.5 depth shows if the adjacent nodes are related to each other, 2 depth presents the direct links of 1 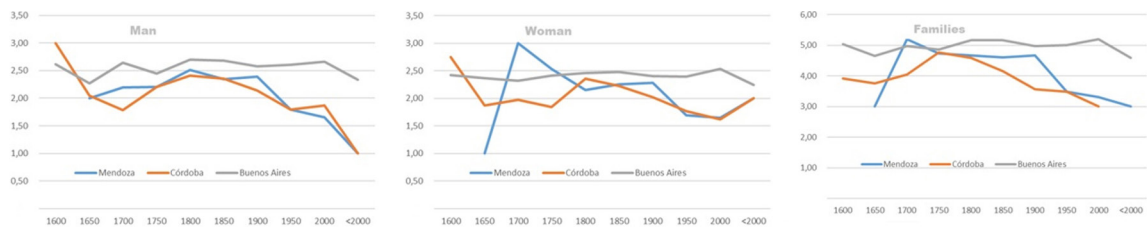depth, and so on. First, the case of H4 (H for *Hombre,* man in Spanish) shows an enormous web of blood and politics after married following his great-great grandson's relations. On the other side, M692279 (M for *Mujer,* woman in Spanish) is a female with a smaller family network and lacks the death date. Finally, M692245 only has a depth ego-net of 1.5, lacking time and geographical information.

Men

Period	Man	Woman	Family
1600	2,62	2,43	5,04
1650	2,27	2,37	4,65
1700	2,64	2,32	4,97
1750	2,45	2,42	4,87
1800	2,70	2,46	5,17
1850	2,68	2,48	5,16
1900	2,58	2,40	4,98
1950	2,61	2,39	5,00
2000	2,66	2,54	5,20
2050	2,33	2,25	4,58

Buenos Aires

Women

Period	Man	Woman	Family
1600	3,00	2,75	5,75
1650	2,04	1,88	3,92
1700	1,78	1,98	3,76
1750	2,21	1,84	4,05
1800	2,41	2,35	4,77
1850	2,36	2,23	4,59
1900	2,14	2,02	4,16
1950	1,80	1,77	3,56
2000	1,87	1,62	3,49
<2000	1,00	2,00	3,00

Córdoba

Family

Period	Man	Woman	Family
1600			
1650	2,00	1,00	3,00
1700	2,20	3,00	5,20
1750	2,20	2,53	4,73
1800	2,52	2,15	4,67
1850	2,35	2,26	4,61
1900	2,39	2,29	4,68
1950	1,79	1,70	3,49
2000	1,65	1,65	3,30
<2000	1,00	2,00	3,00

Mendoza

Table 2. Time gender composition of Buenos Aires, Córdoba, and Mendoza.
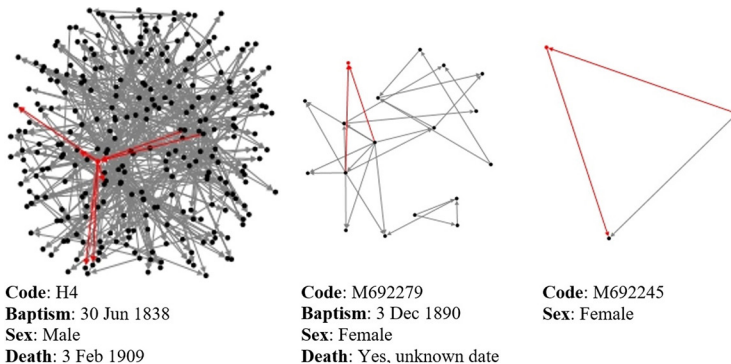
Code: H4
Baptism: 30 Jun 1838
Sex: Male
Death: 3 Feb 1909

Code: M692279
Baptism: 3 Dec 1890
Sex: Female
Death: Yes, unknown date

Code: M692245
Sex: Female

Table 3. Three examples of individuals (4.5 links-depth).

All gathered data is aggregated to kinship social networks sorted by time and space. Figure 5 presents the aggregation for Buenos Aires from 1540 to 2020 in circular layouts. Red nodes represent women, and blue ones represent men. The two periods from 1850 to 1950 seem the biggest, probably in concordance with an increasing population of the country by immigration.
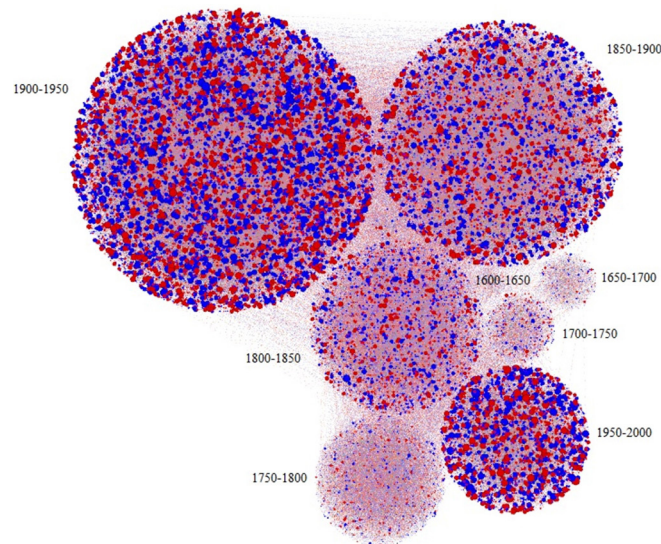
Figure 5. Buenos Aires sorted by 50-year periods (1540-2020).

Figure 6 presents analogous information as Figure 5 but for the province of Mendoza. Again, the giant cluster belongs to the 1850-1950 period. An exciting piece of information derived from genealogical trees is that of lineages. That is a group of individuals tracing descent from a common ancestor. A genealogical network must be acyclic for lineage estimation. An acyclic network is a bipartite graph having no cycles. A connected acyclic network is known as a tree, and a possibly disconnected acyclic network is known as a forest. Lineage estimation requires visiting nodes departing from a common ancestor following down the tree without visiting any previous node. That is the main reason for adopting the family type depicted in Figure 4: it allows for the conformation of an acyclic graph.

Figure 6. Dynamics of the province of Mendoza, 1700-2000.

AWARI
Vol. 3, 1-8, DOI: 10.47909/awari.150

Kinship network evolution in Argentina.
An exploration based on online data

At this initial stage of the project, we only manage to identify and isolate the genealogical trees of Córdoba and Mendoza. Buenos Aires still have cycles. Thus, we cannot provide lineages from the region. We suspect potential endogamy problems in this sample. Figure 7 presents the most extended lineages obtained from the provinces of Córdoba and Mendoza. As observed, Córdoba shows the shortest lineages compared to the other territory.

Córdoba   Mendoza

Figure 7. The longest lineages were estimated from Córdoba and Mendoza.

4. FOLLOWING STEPS

The first step for answering the questions requires information. We choose the kinship network approach. We will form families from records and, mainly, online information. Access to online genealogical information eases the process, with some trade-off in terms of quality.

The project continues by answering different questions related to the influence of family and blood in diverse dimensions of the socio-economic-political life of Argentina. Families played a fundamental role in colonial times in Argentina, as remarked in del Valle y Larrosa (2019). During the Viceroyalty of the River Plate *(Virreinato del Río de la Plata),* merchant families form networks that influence the local political sphere (Figure 1). That influence is known to prevail during the independence process after 1810. How far in time did these families cast a shadow over the economy, politics, and society in the newly formed Argentine Republic? In the broader sense, how do families influence different dimensions of national and regional social life? What role, if any, has kinship played in determining political life? What role, if any, has kinship played in business/commercial life?

We look forward to comparing the composition board of directors in the familiar business. We also are interested in the foundation of new companies and the inheritance of assets (land).

Another question is related to assortativity/homophily. How is this present in the sample? Do these patterns vary over time, with gender, across regions? Another inquiry refers to marriage strategies: can we infer from the data patterns? Are there inbreeding problems?

Family and blood relations come with specific issues. Incest and the consequence of inbreeding problems have always been a topic in the literature (Wright, 1917; van den Berghe, 1983; Shor and Simchai, 2009; Read, 2014). The question will not be treated in immediate contributions but will be on the agenda for future research. Many questions deserve our time and attention shortly.

Contribution statement

Conceptualization, data curation, formal analysis, methodology, validation, visualization, original writing, and review: Juan M.C. Larrosa

Conceptualization, data curation, and revision: Franco Galgano

Revision: Emiliano Gutiérrez

Conflict of interest

The authors declare that there is no conflict of interest.

Statement of data consent

The data generated during the development of this study is still being processed and is not available to third parties. ●

REFERENCES

ALGER, I. & WEIBULL, J. W. (2010). Kinship, Incentives, and Evolution. *American Economic Review, 100:* 1725-1758. DOI: 10.1257/aer.100.4.1725.

BAILEY, D. H., HILL, K. R., & WALKER, R. S. (2014). Fitness consequences of spousal relatedness in 46 small-scale societies. *Biology Letters.* https://doi.org/10.1098/rsbl.2014.0160

COCKERMAN, C. C. (1967). Group Inbreeding and Coancestry. *Genetics, 56,* 89-104. https://doi.org/10.1093/genetics/56.1.89

DEL VALLE, L. C & LARROSA, J. M. C. (2021). La red de familia capitulares de Buenos Aires (1700-1820). Dinámica y ciclos de preeminencia. HiSTOReLo. *Revista de Historia Regional y Local, 13*(28), 96-137. https://doi.org/10.15446/historelo.v13n28.89622

DERRIDA, B., MANRUBIA, S. C., AND ZANETTE, D. H. (1999), "Statistical Properties of Genealogical Trees", *Physical Review Letters 82*(9): 1987-1990. https://doi.org/10.1103/PhysRevLett.82.1987

INDEC (2012). *Censo nacional de población, hogares y viviendas 2010,* Buenos Aires: Instituto Nacional de Estadística y Censos – INDEC.

KOHLER, H. & HAMMEL, E. A. (2001). On the role of families and kinship networks in pre-industrial agricultural societies: An analysis of the 1689 Slavonian census. *Journal of Population Economics, 14,* 21-49. https://www.jstor.org/stable/20007748

MARTÍNEZ ALCALÁ, S.; & ZANETTE, D. H. (2021). Kinship networks in shrinking and growing populations. *Physica A.* https://doi.org/10.1016/j.physa.2021.126554

MURPHY, M. (2011). Long-Term Effects of the Demographic Transition on Family and Kinship Networks in Britain. *Population and Development Review, 37,* 55-80. https://www.jstor.org/stable/41762399

READ, D. W. (2014). Incest Taboos and Kinship: A Biological or a Cultural Story? *Reviews in Anthropology, 43*(2), 150-175, DOI: 10.1080/00938157.2014.903151.

ROTH, C., GARGIULO, F., BRINGÉ, A., & HAMBERGER, K. (2013). Random Alliance Networks. *Social Networks, 35,* 394-405. http://dx.doi.org/10.1016/j.socnet.2013.04.006

SCHULZ, J. F. (2022). Kin Networks and Institutional Development. *The Economic Journal, 132*(647), 2578-2613. https://doi.org/10.1093/ej/ueac027

SHENK, M. K., TOWNER, M. C., VOSS, E. A., & ALAM, N. (2016). Consanguineous Marriage, Kinship Ecology, and Market Transition. *Current Anthropology, 57*(S13). https://doi.org/10.1086/685712

SHOR, E., & SIMCHAI, D. (2009). Incest Avoidance, the Incest Taboo, and Social Cohesion: Revisiting Westermarck and the Case of the Israeli Kibbutzim. *American Journal of Sociology, 114*(6), 1803-1842. http://www.jstor.org/stable/10.1086/597178

WALKER, R. S., & BAILEY, D. H. (2014). Marrying Kin in Small-Scale Societies. *American Journal of Human Biology, 26,* 384-388. doi: 10.1002/ajhb.22527.

WHITE, D. R., & JORION, P. (1996). Kinship networks and discrete structure theory: applications and implications. *Social Networks, 18,* 267-314. https://doi.org/10.1016/0378-8733(95)00277-4

WHITE, D. R.; BATAGELJ, V., & MRVAR, A. (1999). Analyzing Large Kinship and Marriage Networks with Pgraph and Pajek. *Social Science Computer Review, 17*(3), 245-274.

WRIGHT, S. (1922). Coefficients of Inbreeding and Relationship. *The American Naturalist, 51,* 545-559. https://doi.org/10.1086/279872

ZANETTE, D. H. (2019). Statistical properties of model kinship networks, *Journal of Statistical Mechanics: Theory and Mechanics* 094008. https://doi.org/10.1088/1742-5468/ab3720