Key
Vol. no. 1(1) 2025*(June)*: 12-22

Original Article

A comparative analysis of the graph-theoretic representations of the genealogies, a social network approach

Arun Joram[1] and Karam Ratan Singh[2]*

[1]Department of Science and Humanities, Rajiv Gandhi Government Polytechnic College, Itanagar, Arunachal Pradesh, India.

[2]*Department of Basic and Applied Science, National Institute of Technology, Jote-791113, Arunachal Pradesh, India.

***Corresponding author:** karamratan7@gmail.com

Received: 27th September 2025
Accepted: 25th October 2025
Online: 5th November 2025

Abstract

Graph theory and social network analysis provide an excellent framework for visualising and comprehending social networks. A network can be represented by an Ore graph, P graph (or parental graph), or Bipartite P graph, each of which has advantages and disadvantages. The genealogical data on the Galo tribe of Arunachal Pradesh, India, which uses a very distinct naming system to help them remember and recollect all the members of their respective clans, is graphically represented using the three aforementioned graphs to determine the most appropriate representation of genealogical data based on social network parameters. To examine these representations, social network measures are used, and it is discovered that nearly all three representations have roughly comparable centrality ratings. However, the P graph and Bipartite P graph do not demonstrate network transitivity, whereas the Ore graph does with a coefficient of 0.25497421. There were no reconnected marriages among the three representatives.

Keywords: Kinship network; Naming system; Graph theory; Social network analysis

Introduction

The easternmost portion of India's union is located in the Trans-Himalayan region and is known as Arunachal Pradesh, which means "Land of dawn-lit mountains" and is also referred to as the Orchid state of India. Previously known as the North Eastern Frontier Agency (NEFA), this hilly state has a sparse population. The state has 1,382,611 inhabitants and a total land area of 83,743 square kilometres (2011 census of India). The state has 1,382,611 inhabitants and a total land area of 83,743 square kilometres (2011 census of India). With around 110 sub-tribes and 26 major tribes, Arunachal Pradesh is a diversified state in terms of ethnicity. One of the 26 major tribes of Arunachal Pradesh is the Galo (formerly known as the Gallongs), which is distributed throughout the districts of West Siang, Lepa Rada, and Lower Siang. In East Siang, Upper Subansiri, and some areas of Namsai, some population of the tribe can be also be found. The Galo tribe belongs to the Tani group inhabiting Arunachal Pradesh. The naming of kin is a significant part of the Galo tribe's tradition and culture; they have a particularly special technique for naming their children that uses a patrilineal method, which is generally disyllable. The first syllable of the child's name is derived from the last syllable's father's name, (Doye (2020)). For example, if the father's name is *Tomi*, then the names of his children will start from *mi* (last syllable of father's name), like *Migo*, *Mimar*, *Miksen* so on. They distinguish themselves from other tribes in Arunachal Pradesh due to their distinctive naming system. Because the Galo people have been naming their kin in this distinctive pattern for generations, both when naming their different clans and when naming an individual, they can not only trace their ancestry back to Abo Tani (father of the human race) but also helps them to remember and recollect the names of their ancestors from Abo Tani up to their generations. The Galo genealogies do not include females, because of which it has become very difficult to trace back the female descendants of the tribe, (Ete (2021)). Galo people practices both monogamy and polygamy, they perform a variety of marriage rituals and strictly prohibit marriages in the same clan, (Karlo (2017)).

A social network is a social structure that consists of many social actors (such as persons or groups), sets of dyadic links, and other social interactions among the actors. A kinship network is a social network made up of people who are

doi: will be included shortly

related through blood, marriage, or adoption. There are two types of kinship, namely; Consanguineous kinship (based on blood that traces descent) and Affinal kinship (based on marriage, adoption, or other connections).

A kinship network can be defined as a graph $G(V, E, A, \sim)$, where V is a set of vertices that represent individuals, E is a set of edges that signify marriage, and A is a set of arcs referred to as decent arcs in a kinship network, and ~ is an equivalence relation on V that partitions set V into *n* disjoint classes called genders, (Hamberger et al. (2011)). Graph-theoretically, a kinship network can be visualized in two ways, matrix visualization (adjacency matrix) or graph visualization. However, a social network may contain large numbers of nodes and links, so constructing an adjacency matrix for such a network will be very time-consuming and difficult to visualize, because of which graph visualization is always preferred over a matrix (Joram and Singh 2025). We have three visual representations of a social network according to graph theory;

- Ore graph
- P graph or Parental graph
- Bipartite P graph

each has a unique set of benefits and drawbacks.

A collection of techniques for examining the structure of entire social entities are offered by the social network perspective, together with several theories that explain the patterns observed in these structures. Indices such as centrality measures, Freeman (1978) studied the structure of a network at the node level, whereas measures such as graph density (Newman 2018), relinking index (Mrvar and Batagelj 2004), average path length, and clustering coefficients, (Li et al. 2017), (Deepa et al. 2025), examine the structure of the entire network. The present work aims to represent the *Chiram* clan (sub-clan of *Nyochi* group of clans) using the Ore graph, P graph, and Bipartite P graph and to perform a comparative analysis of these representations based on social network parameters.

Data and Methodology

In-person interviews with knowledgeable *Galo* family members of the *Nyochi* group of clans who have preserved their genealogies both in oral and documented formats, as well as research by anthropologists and sociologists, were used to collect the data for this study. Utilizing the Ore graph, P graph, and Bipartite P graph, the network is represented graph theoretically, (Joram et al. (2025). The data has been represented and analyzed using the software, Pajek (Batagelj and Mrvar 2014), and Gephi (Cherven 2015). The analysis of the genealogical data on *Chiram* clan is carried out, based on the following social network metrics.

Centrality measures

Perhaps the most basic social network metrics that can be used to study the structure of a network, are the centrality measures. Centrality is a measure indicating the importance of nodes in the network, it measures their prestige, prominence, importance, and power.

(1) Degree centrality: The degree centrality of a network is given by;

$$C_{deg}(v) = \frac{d(v)}{n-1} \qquad (1)$$

where *d(v)* is the degree of the reference node *v* and *n* denotes the total number of nodes in the network.

(2) Betweenness centrality: It is described as the percentage of a node that is contained inside the other node pairs' shortest pathways. The betweenness centrality is given by;

$$C_{bet}(v) = \sum_{s<t} \frac{b_{st}(v)}{b_{st}} \qquad (2)$$

where $b_{st}(v)$ is all the shortest path that contains node *v* and b_{st} is the shortest path from node *s* to node *t*.

(3) Closeness centrality: It computes the average separation between two vertices. The node that has the highest closeness centrality is the one that is nearest to every other node. It is given by;

$$C_c(u) = \frac{n-1}{\sum_{v=1}^{n} d(u,v)} \qquad (3)$$

Where n is the total number of nodes and d(u,v) is the length of shortest path (geodesic) from node u to node v.

Network density

The ratio of the number of current connections between nodes to the number of potential links is the density of a network.

$$D = \frac{m}{n(n-1)} \qquad (4)$$

The graph has an overall edge count of *m* and an overall node count of *n*.

doi: will be included shortly

Density, which has a range of 0 to 1, depicts how cohesive a network is.

Clustering coefficient

Let *v* be the vertex in a graph *G*. The clustering coefficient for *v* is given by;

$$Cc(v) = \frac{Number\ of\ pairs\ of\ adjacent\ neighbors\ of\ v}{Number\ of\ pairs\ of\ neighbors} \quad (5)$$

The clustering coefficient of a graph *G* is the average of the clustering coefficients of its vertices.

$$Cc(G) = \frac{1}{|V|} \sum_{v \in V} Cc(v) \quad (6)$$

Properties:

- $0 \leq Cc(G) \leq 1$
- $Cc(G) = 1$, implies graph is highly clustered, on average.
- $Cc(G) = 0$, implies graph is not highly clustered, on average.

Relinking Index

It measures the re-linking by marriages among persons belonging to the same ancestry. If *p* denotes the number of vertices, *q* denotes the number of arcs, *l* the number of weakly connected components and *M* the number of maximal (or least) vertices (*outdeg(v)* = 0). Then the Re-linking Index is given by;

$$RI = \frac{l+q-p}{l+p-2M} \quad (7)$$

Properties:

- $0 \leq RI \leq 1$
- $RI = 0$, no re – linking if the network is a tree (forest).

Representation of genealogy

The *Nyochi* group of clans is one of the various clan groups of the Galo tribe, which has maintained its genealogy based on the naming system of kins. The *Nyochi* clan group consists of three sub-clans, namely *Chiram*, *Ropo*, and *Doye*. In the present section, the genealogy of *Chiram* (sub-clan of *Nyochi* clan group) is represented graph-theoretically using the Ore graph, P graph, and Bipartite P graph.

Ore graph

The Ore-graphs, a model of kinship network based on graph theory, was first introduced by *Ostein Ore* in 1960. Male and female individuals are represented by triangles and ellipses, respectively. Marriage is represented by edges, and parent-child relationships are shown by arcs (Hamberger et al. (2011). The Ore graph representation of the *Chiram* clan is shown in figure 1. The network consists of 281 nodes (triangles for males and ellipses for females), 107 edges (blue-coloured) that represent marriages, and 346 arcs that represent parent-sibling relations.

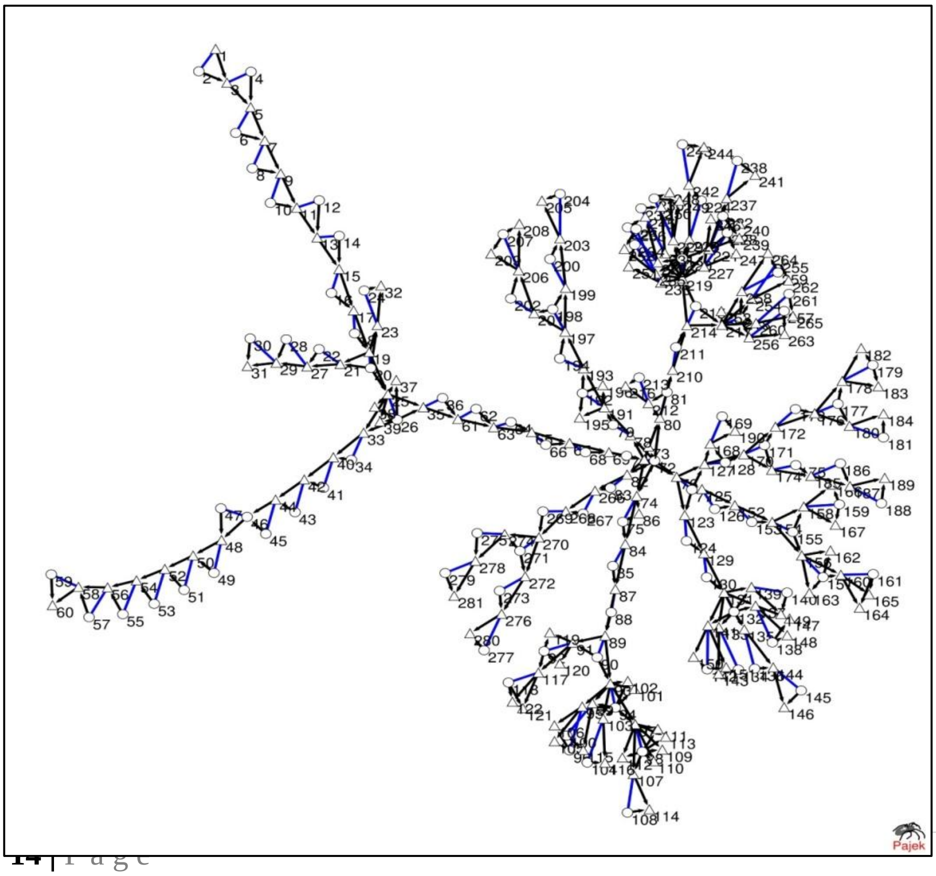

Figure 1. Ore graph representation of genealogy.

doi: will be included shortly

P graph or Parental graph

A kinship network's P-graph representation was first presented by Douglas R. White and Paul Jorion in 1992. P-graphs have vertices that represent married couples or single individuals, with arcs directed downwards from the parents to the children. Males are indicated by solid arcs, and females by dotted arcs (White and Jorion 1992). The P graph representation of the Chiram clan is shown in figure 2, where nodes with distinct colors represent different generations. The underlying P graph of figure 2 is shown in figure 3, where node codes represent individual ids.

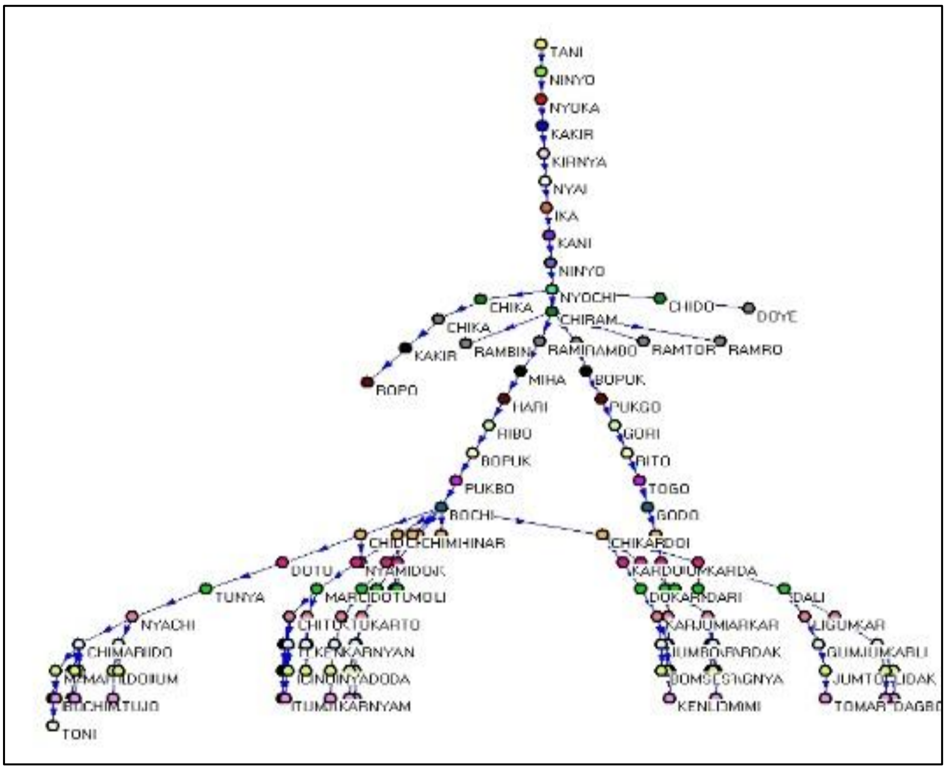

Figure 2. P graph representation of genealogy.

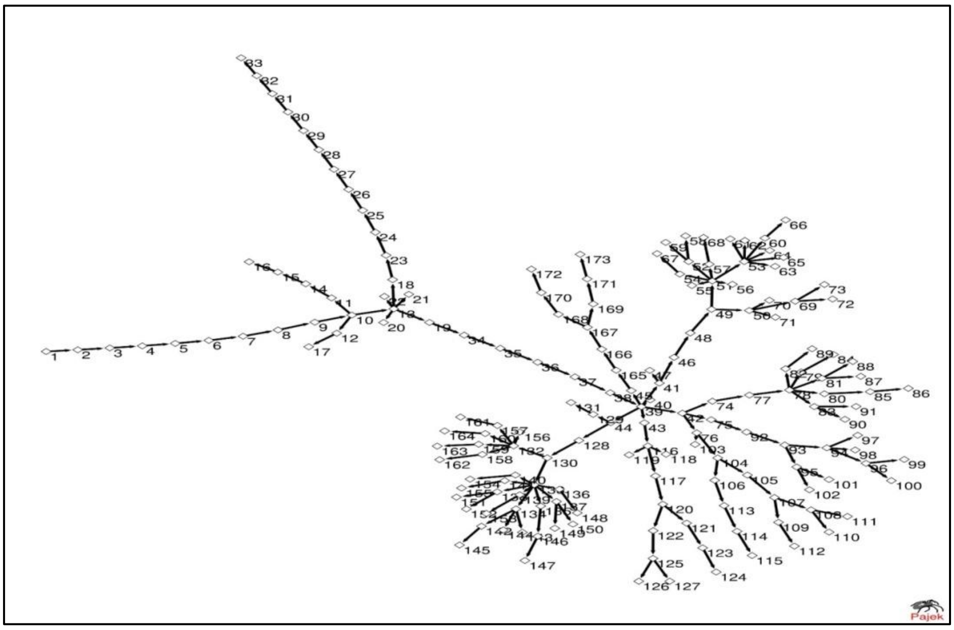

Figure 3. underlying P graph representation of figure 2.

.

doi: will be included shortly

Bipartite P graph

Bipartite P-graphs were first introduced by *Batagelj* and *Douglas R. White* in 2004. Couples (rectangles) and individuals (ellipse for female and triangles for male) are the two different types of vertices in a Bipartite P-graph. As a result, each married person is a part of two different types of vertices. Arcs represent filiations and points from parent to children (Hamberger et al. 2011). The Bipartite P graph representation of the *Chiram* genealogy is shown in figure 4.

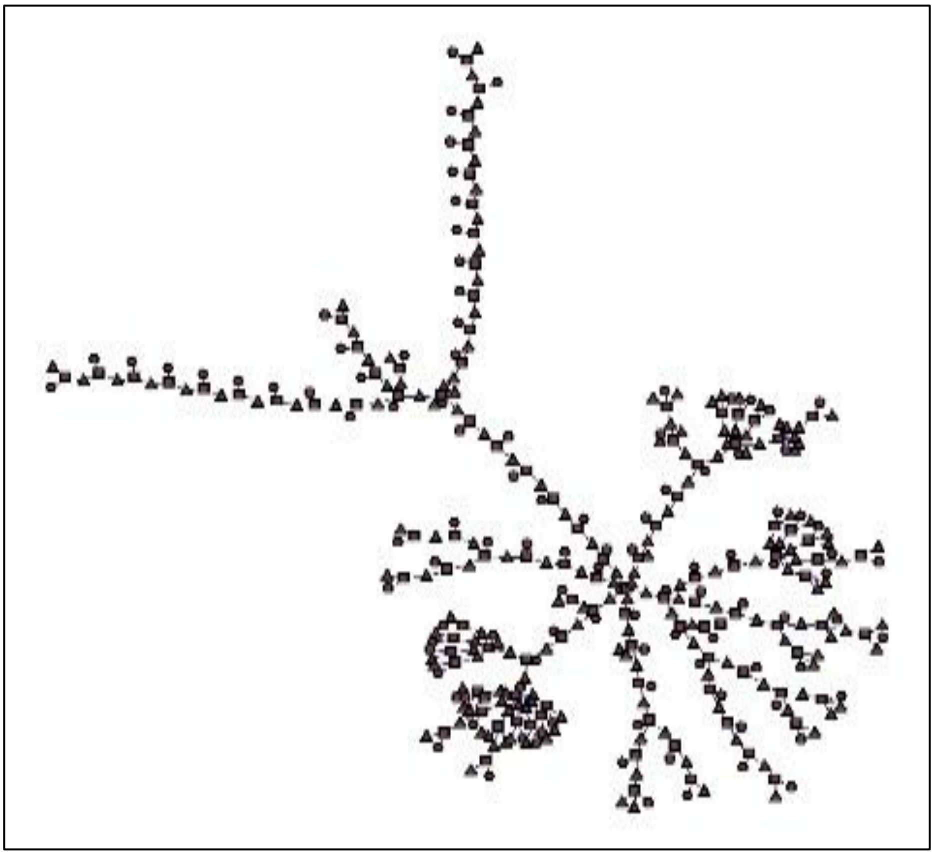

Figure 4. Bipartite P graph representation of genealogy

.

Result and discussion

This section is focused on the analysis of various graph-theoretic representations of the kinship network of Chiram genealogy, based on graph-theoretic parameters. Results obtained on analysis of the kinship network represented by the Ore graph, P graph, and Bipartite P graph respectively shown in table 1.

Table 1. Values of parameters for various graph-theoretic representations.

Graphs	**Nodes**	**Edges /arcs**	**Centrality measures**	**Network density**	**Relinking index (RI)**	**Clustering coefficient**
Ore graph	281	453	$C_{deg}(v) = 9$ $C_{bet}(v) = 0.091398$ $C_c(u) = 0.145333$	0.007	0	0.25497421
P graph	173	172	$C_{deg}(v) = 8$ $C_{bet}(v) = 0.077451$ $C_c(u) = 0.152207$	0.0057	0	0
Bipartite P graph	391	390	$C_{deg}(v) = 8$ $C_{bet}(v) = 0.071268$ $C_c(u) = 0.059772$	0.002557	0	0

doi: will be included shortly

We can observe from table 1 that the P graphs use least number of vertices and edges (arcs) to represent the network in comparison to the Ore graph and Bipartite P graphs. P graph representation of the genealogy consists of 173 vertices that represent a married couple or an unmarried individual, and 172 arcs that descend downward from parent to child, the representation is a tree. Whereas Ore graph representation requires 281 vertices (triangles for males and ellipse for females) and 453 edges (107 edges and 346 arcs), however, it is important to mention here that, the female vertices (ellipse) in Ore graph and Bipartite P graph representations are purely imaginary, this is because the Galo tribes genealogy do not include females. Bipartite P graphs use the highest number of vertices for the representation of the genealogy, this is because, in addition to male and female vertices, an additional vertex (rectangles) is utilized to represent marriage.

Network density

A network's cohesiveness can be ascertained by looking at the density of a graph, which is measured as the ratio of the number of actual ties to potential linkages between nodes. The network density of all the representations is very low, indicating that the network is sparsely connected and that nodes can only communicate with one another via a limited number of pathways. From table 1, the Ore graph representation shows a relatively higher density (0.00709 approx.) than the P graph (0.0057 approx.) and Bipartite P graph (0.00255 approx.). This relatively higher density of Ore graph representation is because of the presence of an edge between male and female nodes (marriage edges).

Centrality measures

(1) Degree centrality: A node's degree, or the quantity of edges it has, is what determines its degree of centrality. The node is more centralized the higher the degree. Both *in-degree(v)* and *out-degree(v)* of vertices in the Ore graph, P graph, and Bipartite P graph representations are determined, however, for the present study, we will consider only the out-degree of nodes, which reflects the number of a male child born to an individual in the clan. The average degree of the network represented by the Ore graph, P graph, and Bipartite P graph is 3.22419929, 1.98843931, and 1.9948849 respectively. In the P graph representation, the individual with the highest degree is *Chitu* with *outdeg(v)*=8, similarly, for the Ore graph, the individual with the highest degree is shared by *Chitu* and the node representing his wife (as arcs descend downwards to siblings from both father and mother), with *outdeg(v)*=9, which is one more than that of P graph, this results because of the arcs that are contributed from the maternal side. However, in-case of Bipartite P graph representation, the highest degree node is the node that represents the marriage between individual *Chitu* and his wife.

Table 2. Degree distribution.

Out-deg(v)	**Ore graph**	**P graph**	**Bipartite P graph**
0	67	67	67
1	0	73	291
2	148	20	20
3	40	6	6
4	12	0	0
5	0	3	3
6	6	3	3
7	6	0	0
8	0	1	1
9	2	0	0

doi: will be included shortly

The higher value of node degree centrality, in the Galo tribe, reflects the number of male siblings born to an individual in the clan, which in turn reflects an individual's importance and prestige in the community. Figure 5, 6 and 7 shows the degree centrality score of the Ore graph, P graph, and Bipartite P graph representations of the clan, respectively. The degree distribution of the Ore graph, P graph, and Bipartite P graph is shown in table 2.

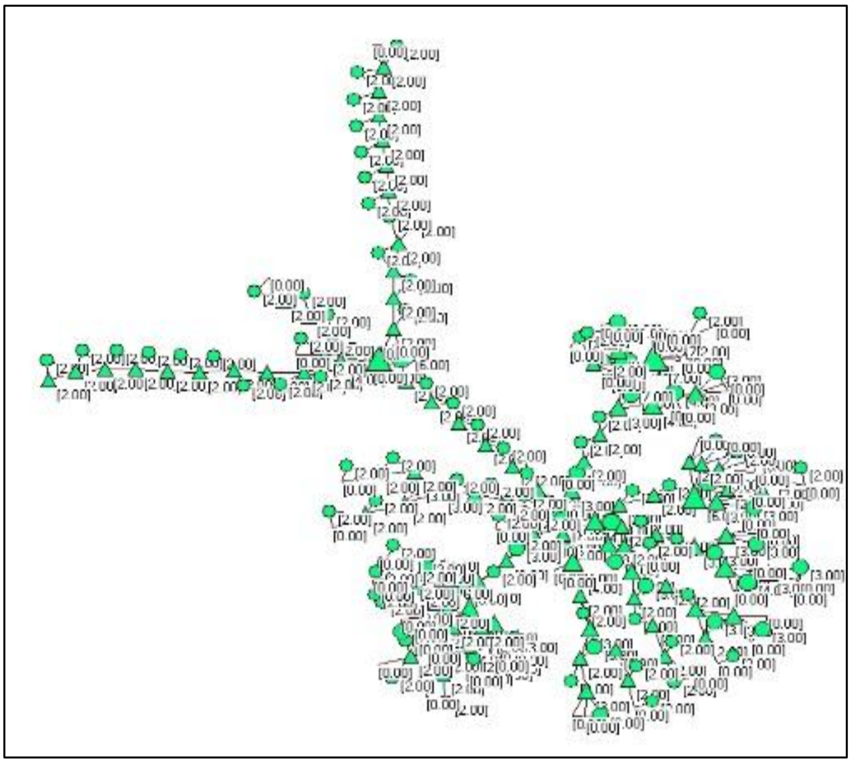

Figure 5. Degree centrality Ore graph.

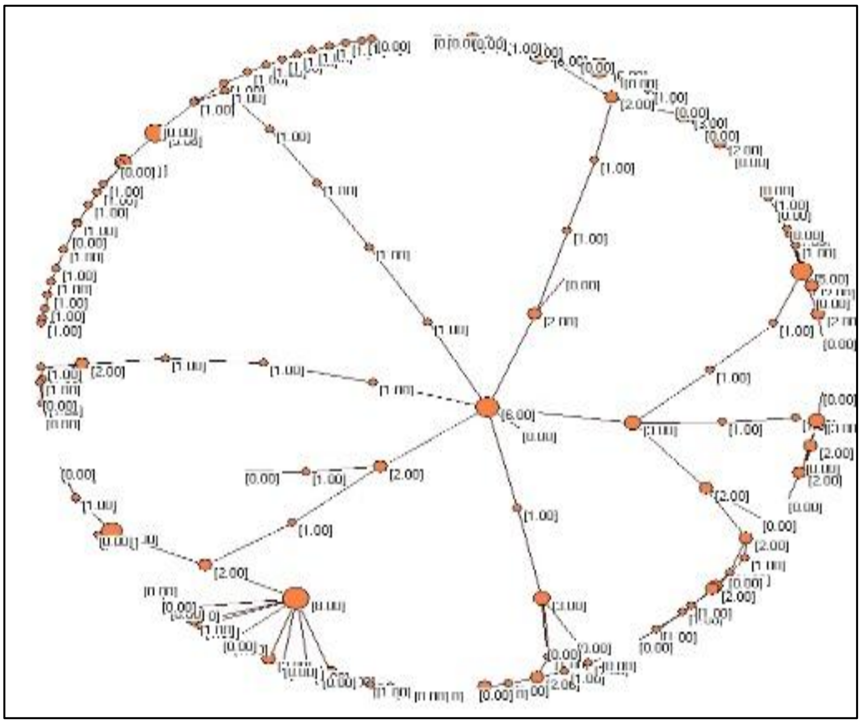

Figure 6. Degree centrality P graph.

doi: will be included shortly

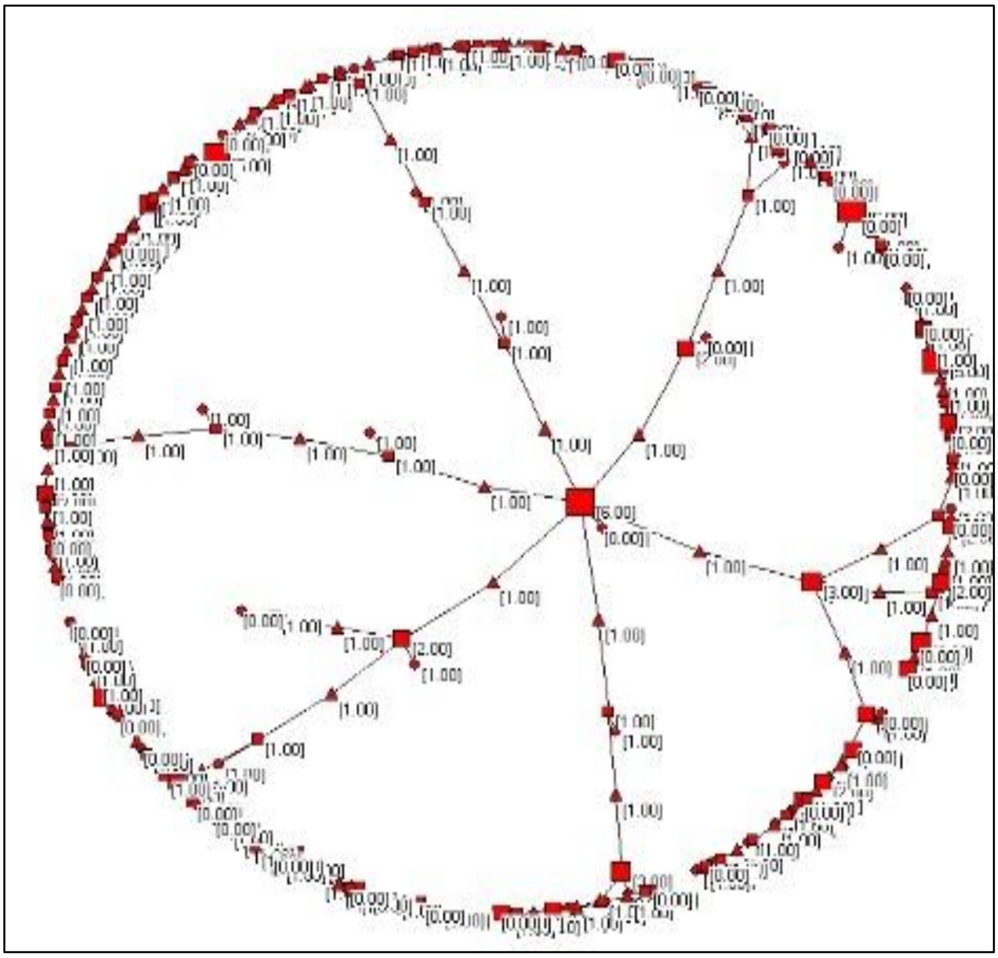

Figure 7. Degree centrality Bipartite P graph.

(2) Closeness centrality: The centrality value, known as closeness centrality measures how far apart nodes are on average from one another. For nodes that are, on average, only briefly separated from one another, closeness centrality takes low values. In social networks, an individual who has a smaller mean distance from other individuals may discover that their viewpoints gain greater community support more quickly than those of other individuals. The highest value of the closeness centrality score for P graph representation is found to be 0.152207, that of individual *Bochi*. However, in the case of Ore graph representation, the highest value of closeness centrality i.e 0.145333 is shared by node *Bochi* and the node that represents his wife, similarly, in the case of Bipartite P graph, the node with the highest value of closeness centrality i.e 0.059772 is the node that represents marriage (rectangles) between *Bochi* and his wife.

(3) Betweenness centrality: The degree to which a node is located on paths that connect to other nodes in the network is measured by betweenness centrality. High betweenness centrality nodes have an impact on the entire network because they regulate the information flow to other nodes. Removal of such nodes from the network disrupts the communication between other nodes and in some cases, these nodes act as cut vertices, whose removal completely disconnects a network.

It is observed that for P graph representation, individual *Bochi* has the highest betweenness centrality of 0.077451. In the case of the Ore graph the highest betweenness centrality score is 0.091398

doi: will be included shortly

(*Bochi*), which is slightly more than that of the P graph, this results because of the female nodes in the network and the marriage edges connecting male and female nodes, which increases the number of possible shortest paths connecting other nodes in the network. However, in the case of Bipartite P graph representation the highest betweenness centrality score is shown by the node reflecting marriage(rectangles) between individual *Bochi* and his wife, with a betweenness centrality score of 0.073713, this is because the arcs descend downwards from both the nodes (triangle for male and ellipse for female) to the marriage node(rectangle).

Relinking Index

Cycles in the P graph reflect relinking marriages; the relinking index evaluates the relinking by marriages among people with the same ancestry (White and Jorion 1992). Galo people firmly forbid Endogamy (marriage in the same clan) and practice Exogamy. Galo people consider it forbidden to marry within the same clan, and those who do so are socially boycotted by society or some penalty is imposed upon the offenders. Galo people engage in both monogamy and polygamy (Karlo (2017)), however, polygamy is not represented in the statistics at hand. The genealogical network of the clan represented by Ore, P, and Bipartite P graphs is found to have a relinking index of 0, as no cycles could be observed in any of the three representations.

Clustering coefficients

The clustering coefficient metric differs from the node centrality measures; it is a metric that is used to measure the transitivity of nodes in the network. Both the global and local (Watts–Strogatz) clustering coefficients of the network are determined for all three representations. It is found that the clustering coefficients (local and global) of the P graph and Bipartite P graph representations are 0, this is because both P graph and Bipartite P graph representations are rooted trees, therefore there are no cliques in the network. However, in the case of the Ore graph representation of the network, the clustering coefficient of the network is found to be 0.25497421. Table 3 shows the local clustering coefficients of the network and the clustering coefficient of the network is shown in figure 8.

Table 3. Clustering coefficients.

Values	Frequency	Frequency %
0.0909	1	0.3559
0.3939	138	49.1103
0.6970	75	26.6904
1.0000	67	23.8434

doi: will be included shortly

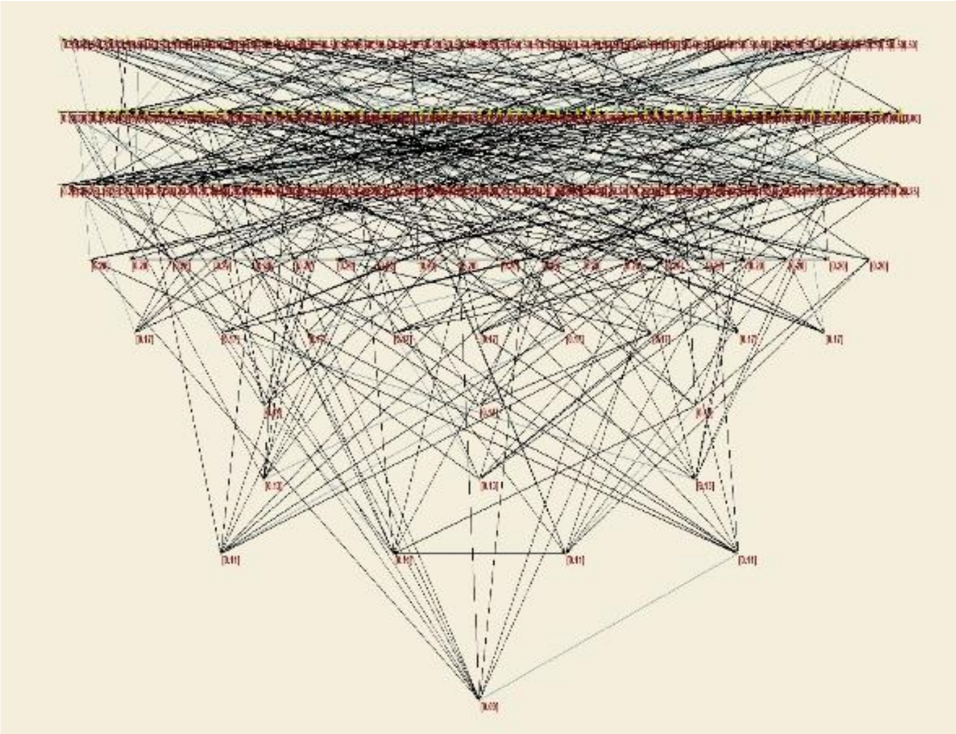

Figure 8. Clustering coefficient of the network

In Ore graph representation, there is an edge between each couple (marriage edge) and arcs descend downwards from both father and mother towards siblings, which results in the formation of triangles or cliques in the network, which in turn results in the detection of clusters in the network. However, Ore graphs are not directed graphs, as it contains both arcs and edges, if we consider the direction of edges into consideration in the network, in such a case no transitivity is shown by the Ore graph as well.

Conclusion

Graph theory provides us with three visual representations of a network namely; Ore graph, P graph, and Bipartite P graph, that can also be utilized to visualize genealogical networks. The present study is carried out on *Chiram* clan, which is a sub-clan of the *Nyochi* group of clans of the Galo tribe of Arunachal Pradesh, India. The Galo tribe's people practice a very unique system of naming their siblings, which helps them to identify, recall, recollect and remember all the members of the clan.

The study aims to select the most suitable visual representation of a genealogical network, based on analysis of the network on social network parameters. P graphs use the least number of vertices and arcs (173 nodes and 172 arcs) to represent the genealogical network, whereas Bipartite P graph representation consists of 391 nodes and 390 arcs and Ore graph uses 281 nodes, 107 edges, and 346 arcs to represent the network. The Galo tribe's genealogy, only includes male members of the clan, it does not reflect any female data or marriages (the tribe practices both monogamy and polygamy), in such a case P graph visualization is more suitable in comparison to the other two. Centrality measures i.e., degree centrality, betweenness centrality, and closeness centrality in the P graph show the centrality score of the nodes or individuals directly involved in the above-mentioned centrality measures and reflects the individual's importance, power, and prestige in the clan. However, in the case of the Ore graph, the highest or lowest centrality score is shared by the node representing an individual (triangle) and the node representing his wife (ellipse), similarly, in the Bipartite P graph representation, the highest or the lowest centrality scores are of the nodes representing marriages.

doi: will be included shortly

P graph and Bipartite P graph networks show no transitivity as the clustering coefficient of these two networks is 0. The network clustering coefficient of Ore graph representation is 0.25497421, however since Ore graphs are mixed graphs (containing edges and arcs) and if we take the direction of arcs into consideration, in such a case the Ore graph network also shows no transitivity. No relinking marriages can be detected in any of the three representations.

Based on the aforementioned findings and analysis, we conclude that P graphs, which employ the fewest nodes and arcs, are more appropriate and practical for representing and visualizing genealogical data. Additionally, P graph representation allows for the more accurate and convenient computation of social network metrics as well as the easier detection of marriage cycles (relinking marriages).

Acknowledgements

We would like to thank *Dr Eli Doye*, Assistant Professor, Government College Tezu, Arunachal Pradesh, for his constant help and timely advice for this work. We are also thankful to all the members of *Nyochi Ao Welfare Society* for sharing data on Nyochi group of clans. Lastly, we would like to acknowledge the support extended to us by *Dr Gomar Basar*, Assistant Registrar, Rajiv Gandhi University Arunachal Pradesh, India.

Declaration of conflict of interest

The authors of this work have no financial or other conflicts to disclose.

Funding

The authors affirm that they did not accept any funds, grants, or other assistance for the creation of this manuscript.

References

Batagelj, V., & Mrvar, A. (2014). *Pajek*, 1, 1245-1256.

Cherven, K. (2015). *Mastering gephi network visualization.* Packt Publishing Ltd.

Deepa, S. N., Singh, K. R., & Joram, A. (2025). Unveiling social network clans: improving genealogical clan classification with SVM neural classifiers and enhanced kernels. *International Journal of Information Technology*, *17*(1), 513-528.

Doye, E. (2020). Understanding naming patterns and kin terms of the galo tribe of Arunachal Pradesh, India. *Ensemble*, 2(1).

Ete, D. (2021). *Missing names. Beyond Myths, a symposium on anthropological histories of tribal worlds*.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239.

Hamberger, K., Houseman, M., and White, D. R. (2011). Kinship network analysis. *The Sage Handbook of Social Network Analysis.*

Joram, A., & Singh, K. R. (2025). *Representation and Analysis of Kinship, Based on the Naming Pattern.* *Journal of Family History*, 50(1), 3-15.

Joram, A., & Singh, K. R. (2025). Comparative study of kinship network community detection approaches. *Discrete Mathematics, Algorithms and Applications*, *17*(03), 2450042.

Joram, A., Ratan Singh, K., Narayan Mishra, L., Rathour, L., & Vanav Kumar, A. (2025). Tracing roots and linkages: Harnessing graph theory and social network analysis in genealogical research, based on the kin naming system. *Discrete Mathematics, Algorithms & Applications*, *17*(4).

Karlo, J. (2017). Etymology of galo marriage system-a brief note on polygamous system. *International Journal of Humanities and Social Science Studies (IJHSSS)*, 4:123–130.

Li, Y., Shang, Y., and Yang, Y. (2017). Clustering coefficients of large networks. *Information Sciences*, 382-383:350–358.

Mrvar, A. & Batagelj, V. (2004). Relinking marriages in genealogies. *Metodoloˇski zvezki-Advances in Methodology and Statistics*, 1:407–418.

Newman, M. (2018). *Networks*, volume 2. Oxford university press.

White, D. R. & Jorion, P. (1992). Representing and computing kinship: A new approach. *Current Anthropology*, 33(4):454–463

doi: will be included shortly