Crossparametric dependencies and Social Network Analysis

Marco Longhin[1], Stefano Ghinoi[*1,2], Gaia Sorge[3], Andrea Artioli[3], and Cristina Guardiano[1,4]

[*]Corresponding Author: stefano.ghinoi@unimore.it
[1]Dipartimento di Comunicazione ed Economia, UNIMORE, Italy
[2]Department of Economics and Management, University of Helsinki, Helsinki, Finland
[3]Dipartimento di Scienze Fisiche, Informatiche e Matematiche, UNIMORE, Italy
[4]Scuola Universitaria Superiore IUSS, Pavia, Italy

The dependency structure of syntactic parameters (Baker, 2001; Guardiano & Longobardi, 2017; Roberts, 2019) poses a challenge for phylogenetic inference, as standard methods (Greenhill et al., 2017) are not designed to accommodate character interdependencies. Building on computational work that represents crossparametric dependencies as network graphs (Kazakov et al., 2018), this study proposes a pilot implementation of Social Network Analysis (SNA) to model parameter systems as relational networks. While the use of SNA is not new in the study of linguistic interactions (Li, Li, & Gao, 2025), its potential for modeling abstract systems of language properties has remained largely unexplored. We pursue two goals: (i) assessing crosslinguistic similarity through network comparison, and (ii) identifying central nodes within parameter networks. Our results show that parameter systems can be characterized using canonical SNA metrics, enabling the quantification of node-level centrality. Systematic node removal yields measurable shifts in similarity scores across language pairs, indicating that parameters contribute asymmetrically to convergence and divergence patterns. These findings suggest that SNA methods may significantly enhance the historical study of syntactic variation by integrating dependency structures into parametric phylogenetic modeling.

1. Deep history and human grammars[1]

Phylogenetic analysis in linguistics has traditionally been grounded in the identification of etymological relations by uncovering systematic regularities underlying morphological structures and sound change (Trask, 1996). Relying predominantly on this type of data, the methods developed by quantitative phylogenetics[2] have achieved unprecedented precision in dating splits and reconstructing migration, contact, and diffusion patterns (Greenhill, 2023).

---

[1] This work is funded by: Next Generation EU, MUR PRIN 2022 project 20224XEE9P (PARTHICO, Grant Assignment Decree No. 901 adopted on 21/06/2023 by the Italian Ministry of University and Research), and by the interdisciplinary project FAR Unimore 2024 Modeling crosslinguistic diversity in Differential Object Marking through the PCM (prot. n. 0323715 del 28/11/2024, Rep. n. 1307/2024). This study was conducted through close collaboration among all authors, who approved the final manuscript. Conceptualization: CG, SG; Methodology and testing: SG, ML, AA; Writing–Original Draft: CG; Writing–Review and Editing: CG, SG, ML, GS.

[2] From at least Gray and Atkinson (2003) up to Heggarty et al. (2023).

Despite this progress, two major aspects remain insufficiently explored: the lack of an uncontroversial method for reconstructing deep linguistic history, and the nature of the historical signal encoded by syntactic structures.

The Parametric Comparison Method (PCM) addresses both issues, positing that "theoretical syntax may provide unexpected evidence for phylogenetic issues typical of the historical-comparative paradigm" (Longobardi & Guardiano, 2009, 1679). This intuition runs counter to centuries of skepticism about the ability of structural patterns to retain historical information, based on the observation that "structural features evolve more rapidly than basic vocabulary [which] is incompatible with hypotheses proposing [a] [. . .] deep signal in grammatical features" (Greenhill et al., 2017). The PCM maintains that, analogously to etymological methods, which identify abstract regularities beyond surface resemblance, the investigation of historical signals in syntax should focus on the abstract combinatorial rules which shape human grammars rather than on surface configurations. Some such rules, encoded in what is technically termed Universal Grammar, while biologically specified and therefore accessible to every language learner, are distributed differently across individual grammars: different grammars may activate different rules. Beginning with Chomsky (1981), these variable rules have been termed *parameters*. Parameters are "points of minimal (binary) choice ultimately responsible for a set of observable differences between two languages" (Crisma, Guardiano, & Longobardi, 2020, 105), called *manifestations*.[3] Given this deductive architecture, languages that activate the same parameter may nonetheless exhibit divergent surface configurations; conversely, identical surface patterns across different languages may result from the interaction of distinct parameters (Guardiano & Longobardi, 2017). Hence, comparisons based on surface patterns are likely to yield substantially different results from those grounded in the underlying rules that generate them. The PCM has shown that language clustering derived from the distribution of parameter values corresponds to established genealogical groupings,[4] while also revealing deeper relationships that lend themselves to formal testing (Ceolin et al., 2021).

2. Parameter systems as dependency networks

Parameter states are not independent to one another.[5] When one state of parameter α entails the irrelevance of parameter β, the manifestations of β become predictable from that state of α; thus, β will not be set at all. This dependency

---

[3] In the PCM, the binary states of parameters are symbolized by [+] and [−]: [+] signals that the parameter is active in the language, i.e. that the language displays at least one of its manifestations; [−] signals that the parameter is absent from the grammar, i.e. none of its manifestations is observed.

[4] Longobardi, Guardiano, Silvestri, Boattini, and Ceolin (2013), Guardiano et al. (2016), Ceolin, Guardiano, Irimia, and Longobardi (2020), Guardiano, Longobardi, Cordoni, and Crisma (2021).

[5] Baker (2001), Fodor (2001), Longobardi (2003), Biberauer and Roberts (2013), Roberts (2019).

structure becomes fully apparent when analyzing a grammatical module of minimally realistic size (Longobardi, 2003), that allows maximizing control over the whole system of dependencies. The PCM formalizes crossparametric dependencies through a list of *implicational conditions*, expressed in Boolean form, either as simple states of other parameters ([+] or [−]) or as conjunctions (written ‘,’), disjunctions (‘or’), or negations (‘¬’) thereof. To provide an example, the implicational condition on parameter TND (*long-distance D-checking demonstratives*) is formulated as [+CGR, (+TSA or +TAR)], meaning that TND can be set if (and only if): (a) parameter CGR (*long-distance Specified Quantity*) is set to [+], and (b) one of the following two conditions holds: (i) TSA (*structural adjectival demonstratives*) is set to [+]; (ii) TAR (*unstructured adjectival demonstratives*) is set to [+]. If, in a language, either condition (a) or condition (b) is not satisfied, TND is neutralized. Neutralization due to a violated condition is encoded as `[0]`.[6]

The pervasiveness of crossparametric dependencies leads to the neutralization of many parameters within each language, resulting in a substantial reduction of the taxonomic characters usable for comparison. Also, it challenges standard automatic models of phylogenetic inference, which are not designed to account for this type of structures (Greenhill, Heggarty, & Gray, 2021). The solution adopted by the PCM has been to employ measures of dissimilarity that ignore neutralized characters or, in character-based analyses, to assign them the status of missing values (Guardiano & Longobardi, 2026). However, both solutions entail loss or distortion of information and are not adequate for measuring the impact of the dependency structure on the diachronic evolution of parameter systems.

Our pilot study explores whether these issues can be addressed using Social Network Analysis (SNA), which enables the mathematical modeling of relational structures and the quantification of network-level properties such as similarity and node centrality (Prell, 2012). We apply these measures to evaluate the contribution of each parameter, modeled as a network node, to overall network organization, by assessing how changes in its states alter the structure and account for variation across networks (i.e. languages).

3. Methods

We use the parameter system described in Crisma, Fabbris, Longobardi, and Guardiano (2025), with minimal adjustments.[7] The dependency structure of parameters is represented as a directed network graph, where parameters are treated as nodes while dependencies define directed edges (Figure 1). Two nodes are connected whenever one parameter entails or neutralises another, the edge capturing

---

[6] For a more detailed description of the type and structure of implicational dependencies in a parametric database, see Guardiano and Longobardi (2017).

[7] The list of the 94 parameters of this dataset and their implicational conditions, as well as all the scripts to reproduce the tests proposed in the present work can be accessed at `https://github.com/marcolonghin/PCM-networks`.

the direction of the dependency. Centrality measures are then employed to quantify the structural prominence of each parameter within the network. We estimate two node-level centrality measures: closeness and betweenness centrality, which measure node accessibility and brokerage, respectively (Prell, 2012). Closeness centrality considers one node's distance to other nodes in the network, while betweenness centrality indicates how many times a node is positioned in the shortest path between two otherwise disconnected nodes; these measures are usually correlated, but they are used for assessing different phenomena. The network structure depends on the capacity of each parameter state to activate the neighbor nodes, assumed to be the main factor driving crosslinguistic differences. In this vein, centrality measures can help in identifying those parameters that are more relevant in activating others and/or working as brokers in developing network chains.

The structure in Figure 1 constitutes the underlying architecture of each language. In the networks representing individual languages, each node is associated to an attribute ([+], [−] or [`0`]). Such networks are then compared to one another to assess their similarity by measuring the role of central nodes in fostering convergence and divergence patterns.

4. Network comparison

We start from a set of six languages: Mandarin (Sinitic); Italian, French, Romanian (Indo-European, Romance); Serbo-Croatian and Bulgarian (Indo-European, Slavic). This selection allows us to check whether the degree of similarity between parameter networks is consistent with previous PCM results, and to test the impact of each node via minimal comparisons between languages at different levels of historical relatedness. We estimate pairwise network similarity using a combination of node attribute distance and edge set distance: the former is defined as the proportion of nodes whose attribute values differ, while the latter is defined as the Jaccard distance between edge sets. Then, we compare the increase (or decrease) in similarity by removing central nodes from the network.

Similarity levels range between 0.681 (Bulgarian vs. Mandarin) and 0.931 (French vs. Italian). All scores are statistically significant at $\alpha = 0.01$. The lowest scores coincide with pairs containing Mandarin, while the highest ones are between languages belonging to one and the same subgroup. This is consistent with the PCM conclusion that the grammars of genealogically related languages are systematically more similar than those of languages belonging to different families (Guardiano et al., 2021). Moreover, the fact that languages known not to be closely related (e.g., Sinitic and Indo-European) yield values near to 0.5 rather than approaching 0 matches the predictions of Guardiano and Longobardi's (2005) *Anti-Babelic Principle*, assessing that, given the constrained nature of parametric variation, genealogically unrelated languages should not display dissimilarity values substantially exceeding those expected by chance. Finally, structural variability across networks aligns with the expectation that genealogically distant (or unrelated) languages activate distinct parameter clusters.

In light of these results, we turn to an analysis that shifts the focus from absolute correlation ranges to the effects of structural manipulation, namely the removal of selected nodes, on pairwise correlations. Figures 2 and 3 show the results of the simulation we use for assessing how languages increase or decrease in similarity by removing central nodes. The main idea behind removing nodes is the following: if similarity increases once a specific node is removed, we assume that the above node is responsible for differentiating two languages; on the other hand, if similarity decreases, it means that the above node is responsible for the convergence between two languages. Our findings show distinct correlation trajectories for closely related versus genealogically unrelated pairs.

The relevant generalizations can be summarized as follows.

(1) All pairs including Mandarin (with the sole exception of Mandarin–Serbo-Croatian) display largely parallel trajectories, regardless of which parameter is removed. This uniform behavior is consistent with the fact that Mandarin has no close historical relationship with any other language in the sample.

(2) Pairs of languages belonging to the same subgroup exhibit parallel trajectories, even when their overall similarity differs. In Figure 2, both Italian–Romanian and French–Romanian increase in similarity when parameters TSA, DOC, and FVP are removed, and decrease when parameters TLC, ACM, and FFP are removed. The same effect is observed in Figure 3, towards the end of the procedure, due to the lower ranking of the relevant parameters in betweeness centrality measurement.

(3) Some marked proximity between Romanian and Bulgarian is observed; however, this effect turns out to be only apparent, as the corresponding curve follows the same trajectory as other Romance–Slavic pairs. In both figures, the consequence of removing parameters NM2 and NUA is that all Romance–Slavic pairs increase in similarity, while, when removing parameters NUC, DGR, NM1, they all decrease.

5. Analytical outlook

These preliminary results suggest that assessing parameter centrality through minimal pairwise network comparisons may provide a predictive tool for estimating the degree of historical relatedness between languages in each pair. Comparisons across pairs further support the prediction that networks representing languages sharing the same degree of historical relatedness exhibit comparable node-centrality configurations. If confirmed by extending the dataset to more diverse and representative language samples spanning different degrees of historical relatedness, these findings are likely to establish network-structure analysis as a viable and theoretically grounded method for quantifying historical relatedness and modeling parameter change.

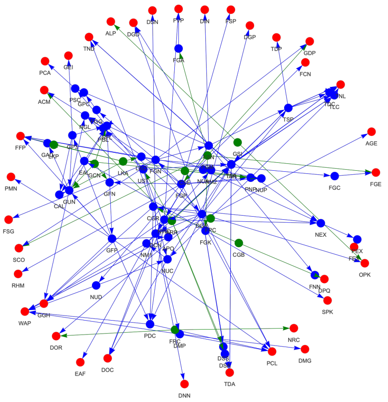

Figure 1. Directed network graph illustrating the dependency structure of the PCM parameter system. Independent parameters (3 of 94) have been removed.

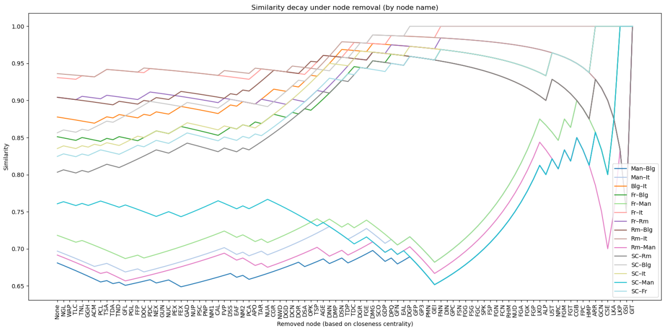

Figure 2. Network similarity and node removal based on closeness centrality

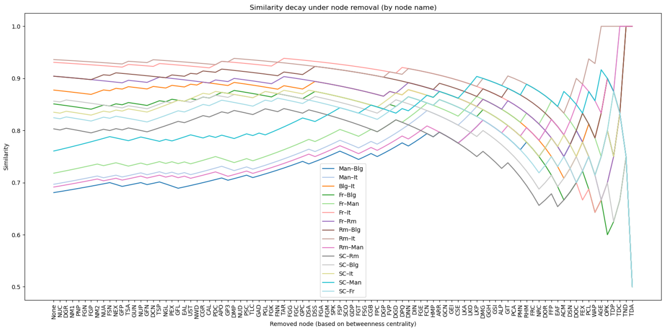

Figure 3. Network similarity and node removal based on betweenness centrality

References

Baker, M. C. (2001). *The atoms of language. The mind's hidden rules of grammar.* Oxford: Oxford University Press.

Biberauer, T., & Roberts, I. (2013). *Emergent parameters. Hierarchies and phylogenies.* (Paper presented at the international workshop Advances in Phylogenetic Linguistics, Ragusa Ibla, Italy, July 15–17, 2013)

Ceolin, A., Guardiano, C., Irimia, M. A., & Longobardi, G. (2020). Formal syntax and deep history. *Frontiers in Psychology*, *11*, 488871.

Ceolin, A., Guardiano, C., Longobardi, G., Irimia, M. A., Bortolussi, L., & Sgarro, A. (2021). At the boundaries of syntactic prehistory. *Philosophical Transactions of the Royal Society B*, *376*, 20200197.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht: Foris.

Crisma, P., Fabbris, G., Longobardi, G., & Guardiano, C. (2025). What are your values? Default and asymmetry in parameter states. *Journal of Historical Syntax*, *9*, 1–26.

Crisma, P., Guardiano, C., & Longobardi, G. (2020). Syntactic diversity and language learnability. *Studi e Saggi Linguistici*, *58*(2), 99–130.

Fodor, J. D. (2001). Setting syntactic parameters. In M. Baltin & C. Collins (Eds.), *The Handbook of Contemporary Syntactic Theory* (pp. 730–767). Oxford: Blackwell.

Gray, R. D., & Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, *426*, 435–439.

Greenhill, S. (2023). Language phylogenies. Modelling the evolution of language. In J. J. Tehrani, J. Kendal, & R. L. Kendal (Eds.), *Oxford Handbook of Cultural Evolution* (pp. 726–744). Oxford: Oxford University Press.

Greenhill, S. J., Heggarty, P., & Gray, R. D. (2021). Bayesian phylolinguistics. In R. D. Janda, B. D. Joseph, & B. S. Vance (Eds.), *The Handbook of Historical Linguistics* (Vol. 2, pp. 226–253). London: John Wiley & Sons.

Greenhill, S. J., Wu, C.-H., Hua, X., Dunn, M., Levinson, S. C., & Gray, R. D. (2017). Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences*, *114*(42), E8822-E8829.

Guardiano, C., & Longobardi, G. (2005). Parametric comparison and language taxonomy. In M. Batllori, M.-L. Hernanz, C. Picallo, & F. Roca (Eds.), *Grammaticalization and Parametric Variation* (pp. 149–174). Oxford: Oxford University Press.

Guardiano, C., & Longobardi, G. (2017). Parameter theory and parametric comparison. In I. Roberts (Ed.), *The Oxford Handbook of Universal Grammar* (pp. 377–398). Oxford: Oxford University Press.

Guardiano, C., & Longobardi, G. (2026). Syntactic methods for language phylogenies. In A. Ledgeway, E. Aldridge, A. Breitbarth, K. É. Kiss, J. Salmons, & A. Simonenko (Eds.), *The Wiley Blackwell Companion to Diachronic and*

Historical Linguistics. Malden: Wiley-Blackwell.

Guardiano, C., Longobardi, G., Cordoni, G., & Crisma, P. (2021). Formal syntax as a phylogenetic method. In R. D. Janda, B. D. Joseph, & B. S. Vance (Eds.), *The Handbook of Historical Linguistics* (Vol. 2, pp. 145–182). Malden: Wiley-Blackwell.

Guardiano, C., Michelioudakis, D., Ceolin, A., Irimia, M., Longobardi, G., Radkevich, N., Silvestri, G., & Sitaridou, I. (2016). South by Southeast. A syntactic approach to Greek and Romance microvariation. *L'Italia dialettale*, *LXXVII*, 95–166.

Heggarty, P., Anderson, C., Scarborough, M., King, B., Bouckaert, R., Jocz, L., Kümmel, M. J., Jügel, T., Irslinger, B., Pooth, R., Liljegren, H., Strand, R. F., Haig, G., Macák, M., Kim, R. I., Anonby, E., Pronk, T., Belyaev, O., Dewey-Findell, T. K., Boutilier, M., Freiberg, C., Tegethoff, R., Serangeli, M., Liosis, N., Stroński, K., Schulte, K., Gupta, G. K., Haak, W., Krause, J., Atkinson, Q. D., Greenhill, S. J., Kühnert, D., & Gray, R. D. (2023). Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages. *Science*, *381*, eabg0818.

Kazakov, D., Cordoni, G., Algahtani, E., Ceolin, A., Irimia, M. A., Kim, S.-S., Michelioudakis, D., Radkevich, N., Guardiano, C., & Longobardi, G. (2018). Learning implicational models of Universal Grammar parameters. In *Proceedings of the 12th International Conference on the Evolution of Language (Evolang12).*

Li, C., Li, Z., & Gao, X. (2025). The application of social network analysis in applied linguistics research. A systematic review. *Applied Linguistics Review*, *16*(4), 1449–1479.

Longobardi, G. (2003). Methods in parametric linguistics and cognitive history. *Linguistic Variation Yearbook*, *3*(1), 101–138.

Longobardi, G., & Guardiano, C. (2009). Evidence for syntax as a signal of historical relatedness. *Lingua*, *119*, 1679–1706.

Longobardi, G., Guardiano, C., Silvestri, G., Boattini, A., & Ceolin, A. (2013). Toward a syntactic phylogeny of modern Indo-European languages. *Journal of Historical Linguistics*, *3*(1), 122–152.

Prell, C. (2012). *Social network analysis. History, theory and methodology.* London: SAGE Publications.

Roberts, I. (2019). *Parameter hierarchies and Universal Grammar.* Oxford: Oxford University Press.

Trask, R. L. (1996). *Historical linguistics. An introduction.* London: Routledge.