



HAL
open science

Reducing networks of ethnographic codes co-occurrence in anthropology

Alberto Cottica, Veronica Davidov, Magdalena Górska, Jan Kubik, Guy
Melançon, Richard Mole, Bruno Pinaud, Wojciech Szymański

► **To cite this version:**

Alberto Cottica, Veronica Davidov, Magdalena Górska, Jan Kubik, Guy Melançon, et al.. Reducing networks of ethnographic codes co-occurrence in anthropology. International Conference on Quantitative Ethnography 2022, Oct 2022, Copenhagen, Denmark. hal-03770039v1

HAL Id: hal-03770039

<https://hal.science/hal-03770039v1>

Submitted on 19 Oct 2022 (v1), last revised 25 Oct 2022 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Reducing networks of ethnographic codes co-occurrence in anthropology [★]

Alberto Cottica^{1[0000-0003-2527-6233]}, Veronica Davidov^{1,2[0000-0002-7098-4338]},
Magdalena Góralaska^{3,4[0000-0001-9491-6682]}, Jan Kubik^{4,5[0000-0003-0017-381X]},
Guy Melançon^{6[0000-0003-3193-7261]}, Richard Mole^{4[0000-0002-4790-5654]}, and
Bruno Pinaud^{6[0000-0003-4814-3273]} Wojciech Szymański^{4[0000-0002-2773-494X]}

¹ Edgeryders

² Monmouth University

³ University of Warsaw

⁴ University College London

⁵ Rutgers University

⁶ University of Bordeaux

Abstract. The use of data and algorithms in the social sciences allows for exciting progress, but also poses epistemological challenges. Operations that appear innocent and purely technical may profoundly influence final results. Researchers working with data can make their process less arbitrary and more accountable by making theoretically grounded methodological choices.

We apply this approach to the problem of reducing networks representing ethnographic corpora. Their nodes represent ethnographic codes, and their edges the co-occurrence of codes in a corpus. We introduce and discuss four techniques to reduce such networks and facilitate visual analysis. We show how the mathematical characteristics of each one are aligned with a specific approach in sociology or anthropology: structuralism and post-structuralism; identifying the central concepts in a discourse; and discovering hegemonic and counter-hegemonic clusters of meaning.

Keywords: anthropology · ethnography · sociology · networks · reduction · visual analytics

1 Introduction

Since their inception, the social sciences have been split between qualitative and quantitative approaches. One of their most challenging undertakings has been to develop multi-method approaches that combine the strengths of both and minimize their weaknesses. We are working on a method that relies on both qualitative and quantitative techniques to increase the benefits of their complementarity. The former are employed at the stage of data collection – via in-depth interviews – and at the stage of analysis, when the ethnographically

[★] Supported by the European Commission’s Horizon 2020 programme, grant agreement 822682.

established contextual knowledge is employed in an iterative interpretation of the collected material in order to reveal repeatable, and thus in some sense "deeper", patterns of thought. Ethnographic coders – who are immersed in the studied societies and cultures – generate rich sets of codes. We analyze them not just to calculate frequencies of themes and motifs, but also to reveal their pattern of connectivity, that we then render in compelling visualizations. In these visualizations, an ethnographic corpus is represented as a network [10], whose nodes correspond to ethnographic codes; the edges connecting them represent the co-occurrence of codes in the same part of the corpus. We call this network a *codes co-occurrence network* (CCN).

A problem that commonly arises is that CCNs are too large and dense for human analysts to process visually. Network science has come up with several algorithms to reduce networks, based on identifying and discarding the least important edges in a network. It is relatively easy to apply them to this type of graph. What is harder is to justify the choice of one or the other of these techniques, and of the values assigned to the tuning parameters that they usually require. In previous work, we have proposed criteria for choosing a technique to reduce a CCN [9], and evaluated four candidate techniques against those criteria. In this paper, we highlight the affinity of each of the four techniques with a prominent method of analysis, associated in turn with a specific school of thought in sociology or anthropology. Next, we use data from a study on Eastern European populism to demonstrate how they work. Our objective is to contribute to the rigor and transparency of the methodological choices of researchers when dealing with large ethnographic corpora.

We proceed as follows. After discussing work related to our own, we introduce the codes co-occurrence network, which is the network to be reduced. Next, we lay out criteria for choosing a technique to reduce a CCN for qualitative analysis, and introduce four such techniques. We then propose a mapping of reduction techniques onto methods of analysis widely used in sociology or anthropology. Finally, we proceed to apply them to our data, to show how the choice of a reduction technique sheds light on a specific facet of the studied phenomena.

2 Related work

The turn towards big data, fueled by improvements in computing power, has led to renewed faith in the ability of quantitative work to provide knowledge that is more generalized than, yet as valid (that is, knowledge that preserves some of the richness of case-derived insights) as that obtainable by qualitative studies or quantitative projects relying on smaller numbers of cases [3].

This has led to exciting progress. At the same time, however, it has highlighted a pressing need for methodological robustness. As scientific work based on large datasets addresses increasingly precise questions, more steps are needed to move from raw data to final result. As a consequence, the methods themselves may be hard to check against the insights derived from intimate familiarity with specific cases. In combination with "publish or perish" and with the premium

placed by journals on counterintuitive, glamorous results, this has led to various epistemological crises. The replication crisis in social psychology is the most famous of them [26], but not the only one. For example, it is claimed that half of the total expenditure on preclinical research in the US goes towards non-replicable studies [14]; and that ostensibly innocent choices about data cleanup prior to analysis might lead to divergent results [11]. Even controlled experiments with different researchers working with the same datasets on the same research questions have led to spectacularly divergent results, for reasons that are not yet entirely clear [32, 5].

Qualitative sociological and anthropological research is not expected to be replicable. Rather, its claim to generating reliable knowledge comes from the rigor and accountability of the methods applied systematically and self-consciously to a specific case or a small range of cases in well-defined spatial and temporal contexts. Therefore, careful, transparent choices about one’s method are necessary every step of the way, even more so when research applies mixed methods [3]. This paper is offered as a contribution to the literature on the significance of such choices in a particular case: that of reducing semantic networks that express qualitative data. The literature on semantic networks originates in computer science [34, 35, 41, 31]; its main idea is to use mathematical objects – graphs – to support human reasoning. Building on this tradition, we focus on the idea of network reduction. In doing so, we factor in previous work on the cognitive limits of humans to correctly infer the topological characteristics of a network from visual inspection [15, 27, 28, 33]. Such work confirms that large and dense networks are hard to process visually, and supports the case for network reduction.

It is important to maintain full awareness of the ways network reduction influences visual interpretation, and to account for them in the analysis. To enhance accountability, we require our mathematical techniques to directly support the specific requirements of knowledge creation in ethnography, and to be intuitive enough to ethnographers. In this sense, this work is inscribed in the tradition of scholars who aim to apply systematic visualization techniques, while still retaining sensitivity to informants’ contextual, interactional, and socioculturally specific understandings of concepts [13, 18, 38, 6].

3 The codes co-occurrence network and its interpretation

Consider an annotated ethnographic corpus. We call any text data encoding the point of view of one informant (interview transcript, field notes, post on an online forum and so on) a *contribution*. Contributions are then coded by one or more ethnographers. Coding consists of associating snippets of the contribution’s text to keywords, called *codes*. The set of all codes in a study constitutes an ontology of the key concepts emerging from the community being observed and pertinent to that study’s research questions ⁷.

⁷ For a complete description of the data generation process, see Section 3 of [10].

We can think of such an annotated corpus as a two-mode network. Nodes are of two types, contributions and codes. By associating a code to a contribution, the ethnographer creates an edge between the respective nodes.

From the two-mode network described above, we induce, by projection, the one-mode CCN. Recall that this is a network where each node represents an ethnographic code. An edge is induced between any two codes for every contribution that is annotated with both those codes (Figure 1). The CCN is undirected ($A \rightarrow B \equiv B \rightarrow A$). There can be more than one edge between each pair of nodes.

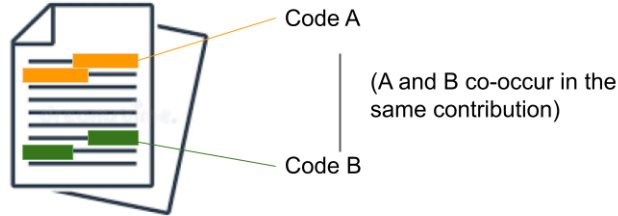


Fig. 1: Inducing a co-occurrence edge between ethnographic codes

We interpret co-occurrence as association. If two codes co-occur, it means that one informant has made references to the concepts or entities described by the codes in the same contribution, seen as a unit. Hence, we assume, both concepts belong to this person's culture-generated mental map. The corpus-wide pattern of co-occurrences is taken to encode the *collective* mental map of informants.

CCNs tend to be large and dense, hence resistant to visual analysis. They are large because a large study is likely to use thousands of codes. They are dense as a result of the interaction of two processes. The first one is ethnographic coding. A rich contribution might be annotated 10 or 20 times, with as many codes associated to it. The second one is the projection from the 2-mode codes-to-contribution network to the 1-mode co-occurrence network. By construction, each contribution gives rise to a complete network of all the codes associated to it, each connected to all the others. Large, dense networks are known to be difficult to interpret by the human eye [15, 27].

4 Techniques for network reduction

Any network reduction entails a loss of information, and has to be regarded as a necessary evil. Reduction methods should always be theoretically founded, and applied as needed, and with caution. We propose four reductions techniques, each one related to a distinct theoretical tradition in the social sciences, particularly anthropology.

Following [20], we propose that a good reduction technique should:

1. Usefully support inference, understood as a simplifying interpretation of the emerging intersubjective picture of the world. The main contribution of network reduction to ethnographic inference is that it makes the CCN small and sparse enough to be processed visually [27, 15]. A well-established literature – and techniques such as layout algorithms – help us define what a “good” network visualization is [19].
2. Reinforce reproducibility and transparency. Reproducibility means that applying the same technique to the same dataset will always produce the same interpretive result (even if the technique has a stochastic component). Transparency means that how the researcher understands how the technique operates, and can explain to her peers how that particular technique contributes to addressing her research question.
3. Not foreclose the possibility of updating via abductive reasoning. Algorithms alone do not decide how parameters should be set to get optimal readability. Rather, the values of the parameters are co-determined by the ethnographers, who possess rich empirical and theoretical knowledge of relevant contexts.
4. Combine harmoniously with other steps of the data processing cycle, such as coding and network construction. This means making sure that the interpretations of the data and their network representation are consistent across the whole cycle.

With that in mind, we turn to the discussion of four candidate techniques. Each of them can be tuned by choosing the value of a reduction parameter (different for each technique) that determines how many edges to discard. The value of this parameter is determined by the researcher, in function of the patterns she explores and of the network topology.

Association depth. A first way to reduce the CCN is the following. For each pair of nodes in the network connected by at least one edge, remove all d edges connecting them, and replace them with one single edge of weight d . This yields a weighted, undirected network with no parallel edges.

d has an intuitive interpretation in the context of ethnographic research. Consider an edge $e = code1 \leftrightarrow code2$. $d(e)$ is the count of the number of times in which $code1$ and $code2$ co-occur. Since we interpret co-occurrence as association, it makes sense to interpret $d(e)$ as the depth of the association encoded in e . This gives us a basis for ranking edges according to the value of d . The higher the value of d of an edge, the more important that edge.

To reduce the network, we choose an integer d^* and drop all edges for which $d(e) \leq d^*$. As the value of d^* increases, so does the degree to which the reduced network encodes high-depth associations between codes.

Association breadth. A second way of reducing the CCN is the following. For all pairs of nodes $code1, code2$ in the network, remove all edges $e : code1 \leftrightarrow code2$ connecting them, and replace them with one single edge of weight b , where b is the number of informants who have authored the contributions underpinning those edges. Like in the previous section, this yields a weighted network of codes with no parallel edges, but now edge weight has a different interpretation: it is a count of the related informants. This has a straightforward interpretation for ethnographic analysis. The greater the value of $b(e : code1 \leftrightarrow code2)$, the more widespread the association between $code1$ and $code$ is in the community that we are studying. We interpret it as association breadth. Notice that $b(e) \leq d(e)$.

As we did for depth, we reduce by choosing an integer b^* and dropping all edges for which $b(e) \leq b^*$. As the value of b^* increases, so does the degree to which the reduced network encodes broadly shared associations between codes.

Highest core values. A third way of reducing the CCN is to consider a co-occurrence edge important if it connects two nodes that are both connected to a large number of other nodes. A community of such nodes can be identified by computing the CCN's k -cores. k -cores are subgraphs that include nodes of degree at least k , where k is an integer. They are used to identify cohesive structures in graphs [16].

After computing all the k -cores of a network, its nodes can be assigned a core value. A node's core value is the highest value of k for which that node is part of a k -core.

To find the most important edges in the CCN, we again replace all edges between any pair of connected codes $code1$ and $code2$ with one single edge $e(code1, code2)$. Next, we choose an integer k^* and remove all the codes c whose core values $k(c) \leq k^*$.

Simmelian backbone. A fourth approach to identify a network's most important edges is to extract its Simmelian backbone. A network's Simmelian backbone is the subset of its edges which display the highest values of a property called redundancy [29]. An edge is redundant if it is part of multiple triangles. The idea is that, if two nodes have many common neighbors, the connection between the two is structural. This method applies best to weighted graphs; in this paper, we use association depth as edge weight.

This technique uses a granularity parameter, k . We set k to be equal to the average degree of the CCN, rounded to the nearest integer. At this point, for each pair of nodes n_1, n_2 , we can compute the redundancy of the incident edge $e(n_1, n_2)$ as the overlap between the k strongest-tied neighbors of n_1 and those of n_2 . To reduce the network, we choose an integer r^* and drop edges for which the redundancy $r(e) \leq r^*$.

5 Mapping network reduction techniques onto four major approaches in sociology and anthropology

Deciding which network reduction technique is best suited to a particular research project depends on the researcher’s ontological and epistemological beliefs, as well as on the nature of the project itself and of its research questions. Each reduction technique reveals a different set of attributes semantic networks have. It also turns out that each technique fits the objectives of a prominent method of analysis, associated in turn with an identifiable approach in sociology or anthropology. Based on this fit, we propose that the researcher’s approach suggests the choice of a reduction technique.

Association depth. Determining association depth is in its essence a method of uncovering the structure of a society or culture. Key works in anthropology – *Anthropologie structurale* [24] and *La Pensée sauvage* [25] – and in social theory [1, 30] initiated a whole host of structuralist and post-structuralist approaches.

For post-structuralist sociologists and anthropologists, social relations can only be understood by analysing how they are constituted and organized through discourse. In other words, social hierarchies, norms and practices are legitimized (or delegitimized) by granting the meaning attached to specific concepts a dominant position, enabling certain ideas to become hegemonic, i.e. widely accepted as then “Truth”. For example, the idea that ethnic nations are natural entities growing out of shared kinship ties (all academic evidence to the contrary) is used to legitimize political control by the core nation and the marginalisation of minority ethnicities. Moreover, discourse scholars work from the assumption that the meaning respondents attach to floating signifiers is relational within a discourse. Within a patriarchal discourse, the meaning attached to ‘woman’ is directly determined by the meaning attached to ‘man’, for instance. To understand the meaning of concepts, it is thus essential to understand their interrelationships; discerning which meanings are hegemonic further requires us to understand which interrelationships between concepts are dominant. Focusing on association depth is thus a useful way of bringing into sharper focus the interrelationships between concepts that are most commonly used by informants, thereby providing a picture of the basic structure of discourse in a given community, within which informants create meaning and make sense of the world around them.

Association breadth. We see association breadth as an alternative point of view on the structure of discourse. Whereas association depth encodes the raw number of co-occurrences between codes, association breadth emphasizes how widespread across different informants those co-occurrences are. In the analysis of section 6 below, we used association depth to check that high-depth edges were not the artefact of just one (or very few) informant who happened to be obsessively associating those particular codes.

Highest core values. The technique based on core values of codes is designed to determine the centrality of certain concepts in a discourse. While it does not allow for the reduction of edges, it shows which concepts have most edges associated with them. It facilitates, therefore, a more systematic determination of which discursive elements constitute what is known in cultural anthropology as root paradigms, key metaphors, dominant schemata or central symbols of a given culture [39, 2].

Simmelian backbone. Finally, the Simmelian backbone extraction can contribute to the discovery of hegemonic and counter-hegemonic clusters (subcultures) of meaning in an analyzed body of discourse [17, 22]. No society or culture is fully integrated and each is subjected to centripetal and centrifugal forces simultaneously. As a result, even in the most “homogenous” societies and cultures one can identify at least embryonic subcultures or – in another formulation – for every hegemony there is a budding or fully articulated counter-hegemony. The point is that a hegemony or counter-hegemony is usually built not on a single symbol or concept but on their interconnected cluster. This reduction technique helps to identify such clusters and assess with greater precision their shape and internal coherence.

6 An application

We used the corpus of a project we are working on to show how each of the four aforementioned reduction techniques can be seen as broadly corresponding to a paradigm in anthropology – a convergence that attests to the utility of such a synthesis. This application is not meant as a full methodological primer. Rather, it means to be a “proof of concept”, and show the possibilities of synthesizing quantitative and qualitative techniques in the service of ethnographic insight.

The data were gathered in the spring and summer of 2021, as a part of a larger research project on populism in Central and Eastern Europe, to be completed by the end of 2022. They consist of 17 semi-structured interviews with Polish-speaking Internet users, who used social media to seek and share information about health against the backdrop of the COVID-19 pandemic. Research participants were asked about their opinion on the current state of affairs in their respective countries, and their political choices over the years and at present. The interviews’ transcriptions (about 78,000 words) were then split into contributions, in the sense of Section 3: each question of the interviewer, and answer of the interviewee was considered as a contribution. In what follows, two codes are considered to co-occur if, and only if, they were both used in annotating the same contribution (as opposed to the same interview). Computed this way, the CCN from this corpus includes 1,116 contributions, and 2,152 annotations. The latter use 600 unique codes, connected by 16,370 co-occurrence edges.

We apply reduction techniques to the CCN in sequence, trying for different levels of the respective reduction parameters (d , b , k , r) in order to achieve a good combination of legibility (more edges discarded) and completeness (fewer

edges discarded). In each reduced network, we focus on the ego network of one code in particular, *Catholic Church*. Ego network analysis is widely used in anthropology, for example in the conventions of kinship charts. We selected this particular code in the expectation that the Catholic Church would be fairly central in any ethnographic study of populism in Poland, and that, therefore, it would appear in most reduced networks.

Highest core values. Anthropology as a discipline has a long history of trying to identify “core” dimensions of culture, both to better theorize how a given culture is constituted, and as a useful heuristic for ethnographic fieldwork (cf Boas’s outer and inner forces [4], Kroeber’s reality and value culture [21], Steward’s cultural core [37]). In our approach we are particularly inspired by Victor Turner, a founding figure in symbolic anthropology – a theoretical approach in British anthropology arising in the 1960s – that viewed culture as an independent system of meaning deciphered by interpreting key symbols and rituals [36] and theorized that “beliefs, however unintelligible, become comprehensible when understood as part of a cultural system of meaning” [12]. Turner subscribed to a definition of symbol as “a thing regarded by general consent as naturally typifying or representing or recalling something by possession of analogous qualities or by association in fact or thought” [40]. As we are invested in holistically understanding and visualizing how cultural beliefs and discourses are assembled, it is the recollection and association aspects that are of particular interest to us. Turner did not seek to define a fixed core of concepts within a culture the way Steward, for example, did. Nevertheless, he did write about symbols “variously known as ‘dominant,’ ‘core,’ ‘key,’ ‘master,’ ‘focal,’ ‘pivotal,’ or ‘central’ [that] constitute semantic systems in their own right [with a] complex and ramifying series of associations as modes of signification.”

We envision network reduction based on the highest core values as revealing something akin to such a semantic system. We approach it in the spirit of Turner’s notion of “positional meaning” articulated in his methodology for studying rituals – a level of symbolic meaning derived from analyzing a symbol’s association to other symbols and cultural concepts, in other words, contextual meaning: “The positional meaning of a symbol derives from its relationship to other symbols in totality, a *Gestalt* whose elements acquire their significance from the system as a whole. This level of meaning is directly related to the important property of ritual symbols... their polysemy. Such symbols possess many senses, but contextually it may be to stress one or a few of them only.” [40]

In our data, we see the highest core values reduction yielding an innermost nucleus of nodes (ethnographic codes) that recur most often in relation with each other. *Catholic Church* is close to the center of the symbols expressing this culture. Mathematically, it belongs to one of the innermost k -cores, ($k = 28$, containing 82 codes, shown in figure 2), though not the absolute innermost. Two k -cores exist in the graph where k is higher than 28 ($k = 29$, $k = 42$). The analysis supports the conclusion that the Catholic Church is one of the core symbols in this culture.

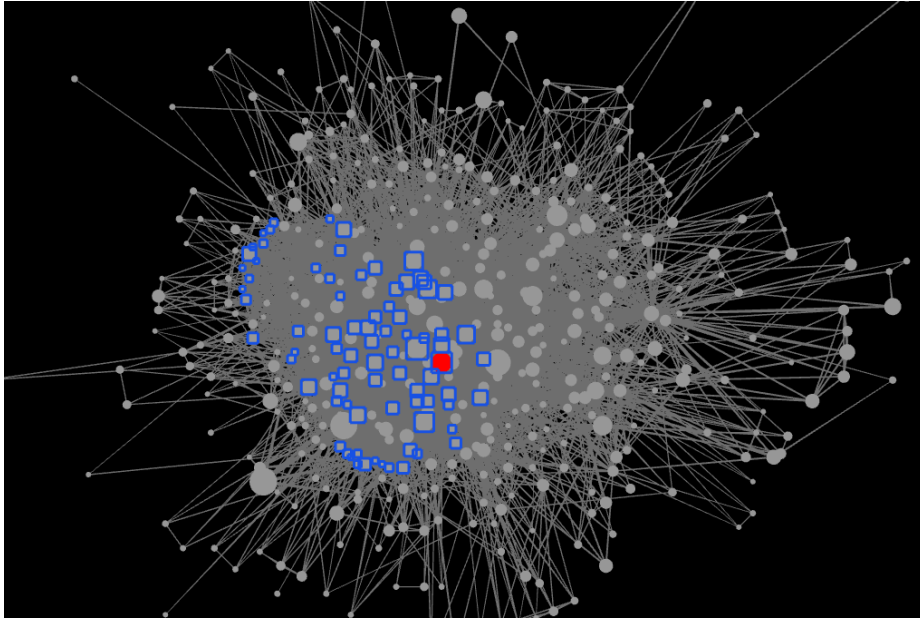


Fig. 2: The full CCN. The 28-core is shown highlighted in blue. It contains Catholic Church (in red).

Simmelian backbone. Next, we explore the neighborhood of `Catholic Church` through the lens of the Simmelian backbone reduction technique. Recall that this technique detects community of nodes connected by redundant links, and was developed to identify homophily and strong ties in a social network of actors [29]. Here, we use it to identify communities of ethnographic codes. In a way, when applied to concepts rather than human actors, this approach literalizes the notion of certain ideas being “in conversation” with each other. The visualization reveals several such “conversations”. The community structure itself maps onto the anthropological notion of culture as a field of competing forces, with different clusters of codes encoding different strands of culture. In the words of Jean and John Comaroff, “culture [is] the semantic space, the field of signs and practices, in which human beings construct and represent themselves and others, and hence their societies and histories. . . culture always contains within it polyvalent, potentially contestable messages, images, and actions.” [8] This approach stresses that culture is neither monolithic nor fixed, but rather always contingent and in flux, and allows us to see, from a bird’s eye perspective, how various “signifiers-in-action” coalesce into identifiable semantic subspaces.

`Catholic Church` belongs to a community of codes that are political rather than spiritual— such as `abuse of power`, `political marketing`, and `right wing` (figure 3. In fact, the highest-redundancy edge incident to `Catholic Church` is to `politicisation` ($r = 55$). Our ethnographic interpretation is that people

have concerns pertaining to the Catholic Church, both in the context of what they conceive as this institution's excessive politicization and more personal concerns, anxieties, and anomic tendencies. This can be used as a foundation to build on iteratively in future research on a range of subjects, including but not limited to political cultures, epistemologies, various dimensions of trust and belief, and the position of the Catholic Church in the public space and the country's culture.

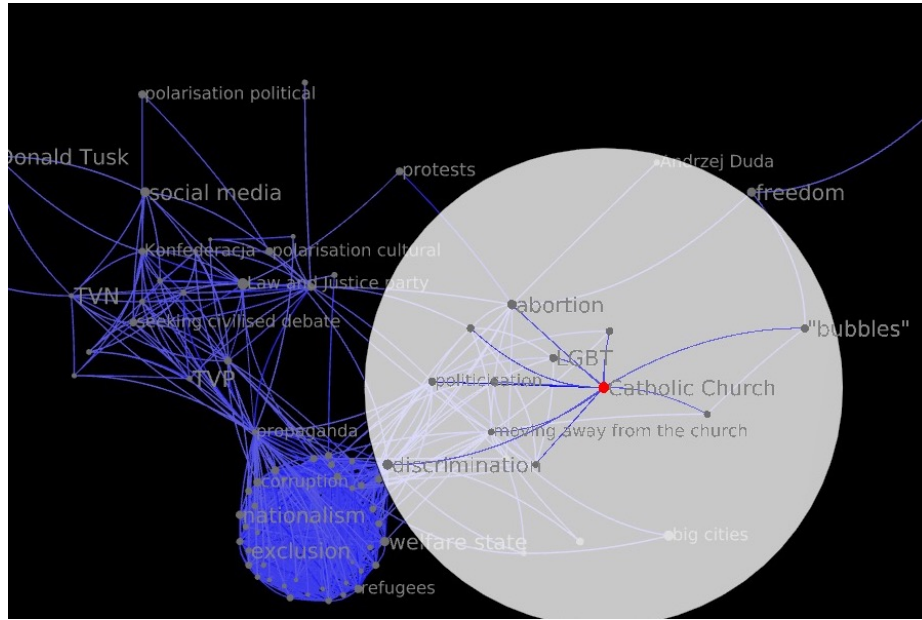
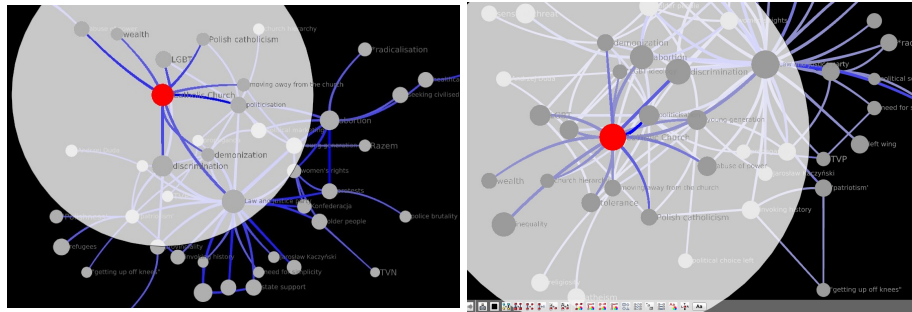


Fig. 3: The ego network of Catholic Church, with only edges with edge redundancy $r > 30$ shown.

Association depth and association breadth. We now turn to the association depth and association breadth reduction techniques, which work in tandem to deepen our understanding of the underlying structures of discursive associations. The association depth visualization shows us which associative links between concepts are the strongest – in other words, which codes emerge as being mentioned together most often. Association breadth helps evaluate the diffusion of these “deep” edges among informants. When the results produced through the depth and breadth reductions align, it confirms that deep associations are not generated by a small number of interviews with people who frame a topic by linking it repetitively with a constant, limited set of other topics, but rather a broad agreement that emerges from the analysis of many interviews or conversations. We can see how this plays out with Catholic Church code (Figure 4a): the three deepest

associations are formed between it and the abuse of power, politicisation, and Polish catholicism codes. If we choose lower (but still significant, in the sense that the number of edges in the CCN is reduced by over 95%) levels of the reduction parameter d , codes like LGBT, discrimination and Law and Justice party appear.

The association breadth-reduced CCN shows that the broadest links to Catholic Church are very similar to the deepest ones. The very broadest three connect it to politicization, Polish catholicism, and discrimination, and tolerance. Edges to “political” codes like LGBT, inequality, abuse of power and abortion resist to reductions by over 50% in the number of edges in the CCN (Figure 4b). In our case, these two reduction methods yield closely aligned results. Both attest to the Catholic Church figuring as an institution associated with politics more so than with faith or spirituality among the informants. Even though there are some codes visible in the graph that may correlate to spirituality, the broadest associations still link the Catholic Church with political codes and the issue of abuse of power.



(a) Only edges with association depth $d > 4$, and incident codes are shown.

(b) Only edges with association breadth $b > 3$, and incident codes are shown.

Fig. 4: The ego network of Catholic Church in two CCN reductions.

7 Discussion and conclusions

As ethnographers working with this form of data analysis, we look for patterns that are of interest to us either for their novelty (unexpected connections) or confirmation of either previous research, or initial impressions formed during data collection. In this particular example, this finding aligns with existing survey studies on Polish Catholicism today, which show that most Poles disapprove of the Church’s direct involvement in politics [7].

More broadly, this approach is synergetic with anthropology’s long-standing interest in structures. While we don’t aspire to resurrect the classic structuralist goal of uncovering deep underlying structures or cross-cultural universals à la

Claude Lévi-Strauss, *the figurehead of structural anthropology* there is methodological value in understanding cultural structures in a way aligned with schema theory developed by cognitive anthropologists rather than old-school structuralists. We are aware that such structures are historically contingent and subject to change. Nevertheless, these visualizations offer us a synchronic snapshot of how people mentally organize their experiences and understanding. From a methodological standpoint this can be valuable not only as new insight or confirmation, but also as a part of an iterative research process. Once we have a sense of what ideas the people under study believe link together most strongly, that knowledge can inform subsequent questionnaires, interviews, and selection of sites for participant observation. For example, perhaps the most salient participant observation in an ethnographic project on the Catholic Church in Poland today would have to take place, counterintuitively, outside the churches, in the domain of politics.

Lévi-Strauss believed that universal deep cognitive structures underpin all human cultural experience; in that, he exemplified the cross-cultural universalism position in anthropology. We do not subscribe to such a position; nevertheless, his approach to myth analysis resonates with our reduction techniques. According to him, all existing versions of the myth had to be aggregated, so that one could isolate what he called “gross constituent units” – clusters of specific types of relations that are present in all versions of the myths (e.g. characters overrating kinship relations, characters underrating kinship relations) [23]. These units, Lévi-Strauss posited, revealed deep structures expressed through the language of myth. In a similar way, we also look at the highest-redundancy edges in order to glean what they reveal about deep associations structuring cultural discourses in a corpus.

In conclusion, through this demonstration we aim to make a contribution to the ongoing and worthwhile conversations in the social sciences geared at synthesizing qualitative and quantitative methods. The reduction techniques discussed in this paper can be instrumental in supporting ethnographic insights, and the accountability of methodological choices in ethnographic research. The ethnographer’s goal and research question inform the choice of a reduction technique; the appropriateness of such choice can be transparently argued by the researcher. Moreover, since the steps to build and reduce the CCN are reproducible (given the value of the reduction parameter), other researchers can validate, dispute, or improve upon her interpretation, thereby contributing to the accountability of qualitative research. The highest core values reduction identifies concepts of central significance, and can help map a starting point of entry into the data; the Simmelian backbone reduction maps heterogeneous communities of meaning, and may be especially helpful in identifying hegemonic and counter-hegemonic discourses at work within a community. Finally, the association depth and association breadth reductions, working in tandem, can help illuminate and validate the most significant associative structures of meaning in specific domains within a community under study.

References

1. Althusser, L.: *For Marx*. Verso Books (1965)
2. Aronoff, M.J., Kubik, J.: *Anthropology and political science: A convergent approach*, vol. 3. Berghahn Books (2013)
3. Beaulieu, A., Leonelli, S.: *Data and Society: A Critical Introduction*. SAGE (2021)
4. Boas, F.: The aims of anthropological research. *Science* **76**(1983), 605–613 (1932)
5. Breznau, N., Rinke, E.M., Wuttke, A., Adem, M., Adriaans, J., Alvarez-Benjumea, A., Andersen, H.K., Auer, D., Azevedo, F., Bahnsen, O., et al.: Observing many researchers using the same data and hypothesis reveals a hidden universe of data analysis. *MetaArXiv* (2021). <https://doi.org/10.31222/osf.io/cd5j9>
6. Burrell, J.: The field site as a network: A strategy for locating ethnographic research. *Field Methods* **2**(21), 181–199 (2009)
7. CBOS Foundation: *Postawy wobec obecności religii i kościoła w przestrzeni publicznej* (2022), https://www.cbos.pl/SPISKOM.POL/2022/K_003_22.PDF
8. Comaroff, J., Comaroff, J.: *Ethnography and the historical imagination*. Routledge (2019)
9. Cottica, A., Hassoun, A., Kubik, J., Melançon, G., Mole, R., Pinaud, B., Renoust, B.: Comparing techniques to reduce networks of ethnographic codes co-occurrence (Jul 2021). <https://doi.org/10.5281/zenodo.5801464>
10. Cottica, A., Hassoun, A., Manca, M., Vallet, J., Melançon, G.: Semantic social networks: A mixed methods approach to digital ethnography. *Field Methods* **32**(3), 274–290 (2020)
11. Decuyper, A., Browet, A., Traag, V., Blondel, V.D., Delvenne, J.C.: Clean up or mess up: the effect of sampling biases on measurements of degree distributions in mobile phone datasets. *arXiv preprint arXiv:1609.09413* (2016)
12. Des Chene, M.: *Symbolic anthropology*. In: *Encyclopedia of Cultural Anthropology*. Henry Holt (1996)
13. Dressler, W.W., Borges, C.D., Balierio, M.C., dos Santos, J.E.: Measuring cultural consonance: Examples with special reference to measurement theory in anthropology. *Field Methods* **17**(4), 331–355 (2005)
14. Freedman, L.P., Cockburn, I.M., Simcoe, T.S.: The economics of reproducibility in preclinical research. *PLoS biology* **13**(6), e1002165 (2015)
15. Ghoniem, M., Fekete, J.D., Castagliola, P.: On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis. *Information Visualization* **4**(2), 114–135 (2005)
16. Giatsidis, C., Thilikos, D.M., Vazirgiannis, M.: Evaluating cooperation in communities with the k-core structure. In: *2011 International conference on advances in social networks analysis and mining*. pp. 87–93. IEEE (2011)
17. Gramsci, A.: *I quaderni del carcere*. Einaudi (1975)
18. Hanmerz, U.: The global ecumene as a network of networks. In: Kuper, A. (ed.) *Conceptualizing society*, pp. 34–56. Routledge (1992)
19. Herman, I., Melancon, G., Marshall, M.: Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics* **6**(1), 24–43 (2000). <https://doi.org/10.1109/2945.841119>
20. King, G., Keohane, R.O., Verba, S.: *Designing social inquiry: Scientific inference in qualitative research*. Princeton university press (1994)
21. Kroeber, A.: Reality culture and value culture. In: *SCIENCE*. vol. 111, pp. 456–457. AMER ASSOC ADVANCEMENT SCIENCE 1200 NEW YORK AVE, NW, WASHINGTON, DC 20005 (1950)

22. Laitin, D.D.: *Hegemony and culture: Politics and change among the Yoruba*. University of Chicago Press (1986)
23. Lévi-Strauss, C.: The structural study of myth. *The journal of American folklore* **68**(270), 428–444 (1955)
24. Lévi-Strauss, C., Lévi-Strauss, C.: *Anthropologie structurale*, vol. 171. Plon Paris (1958)
25. Lévi-Strauss, C., et al.: *La pensée sauvage*, vol. 289. Plon Paris (1962)
26. Maxwell, S.E., Lau, M.Y., Howard, G.S.: Is psychology suffering from a replication crisis? what does “failure to replicate” really mean? *American Psychologist* **70**(6), 487 (2015)
27. Melançon, G.: Just how dense are dense graphs in the real world? a methodological note. In: *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*. pp. 1–7 (2006)
28. Munzner, T.: *Visualization analysis and design*. CRC press (2014)
29. Nick, B., Lee, C., Cunningham, P., Brandes, U.: Simmelian backbones: Amplifying hidden homophily in facebook networks. In: *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on*. pp. 525–532 (Aug 2013)
30. Poulantzas, N.: *On social classes*. *New Left Review* (1973)
31. Shapiro, S.C.: Representing and locating deduction rules in a semantic network. *ACM SIGART Bulletin* (1977). <https://doi.org/10.1145/1045343.1045350>
32. Silberzahn, R., Uhlmann, E.L., Martin, D.P., Anselmi, P., Aust, F., Awtrey, E., Bahnik, Š., Bai, F., Bannard, C., Bonnier, E., et al.: Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science* **1**(3), 337–356 (2018)
33. Soni, U., Lu, Y., Hansen, B., Purchase, H.C., Kobourov, S., Maciejewski, R.: The perception of graph properties in graph layouts. *Computer Graphics Forum* **37**(3), 169–181 (2018). <https://doi.org/10.1111/cgf.13410>
34. Sowa, J.F.: *Conceptual structures: information processing in mind and machine*. Addison-Wesley Pub., Reading, MA (1983)
35. Sowa, J.F., et al.: *Knowledge representation: logical, philosophical, and computational foundations*, vol. 13. Brooks/Cole Pacific Grove, CA (2000)
36. Spencer, J.: *Symbolic anthropology*. In: *Encyclopedia of Social and Cultural Anthropology*. Henry Holt (1996)
37. Steward, J.H.: *Theory of culture change: The methodology of multilineal evolution*. University of Illinois Press (1972)
38. Strathern, M.: Cutting the network. *The Journal of the Royal Anthropological Institute* **2**(3), 517–535 (1996)
39. Turner, V.: Liminal to liminoid, in play, flow, and ritual: An essay in comparative symbolism. *Rice Institute Pamphlet-Rice University Studies* **60**(3) (1974)
40. Turner, V.: Symbolic studies. *Annual Review of Anthropology* **4**(1), 145–161 (1975)
41. Woods, W.A.: What’s in a link: Foundations for semantic networks. In: *Representation and understanding: Studies in Cognitive Science*, pp. 35–82. Elsevier (1975)