Page Proof Instructions and Queries

Journal Title: FMX

Article Number: 908236

Thank you for choosing to publish with us. This is your final opportunity to ensure your article will be accurate at publication. Please review your proof carefully and respond to the queries using the circled tools in the image below, which are available in Adobe Reader DC* by clicking **Tools** from the top menu, then clicking **Comment**.

Please use ***only*** the tools circled in the image, as edits via other tools/methods can be lost during file conversion. For comments, questions, or formatting requests, please use T. Please do ***not*** use comment bubbles/sticky notes.

*If you do not see these tools, please ensure you have opened this file with **Adobe Reader DC**, available for free at https://get.adobe.com/reader or by going to Help > Check for Updates within other versions of Reader. For more detailed instructions, please see https://us.sagepub.com/ReaderXProofs.

No.	Query
	Please confirm that all author information, including names, affiliations, sequence, and contact details, is correct.
	Please review the entire document for typographical errors, mathematical errors, and any other necessary corrections; check headings, tables, and figures.
	Please confirm that the Funding and Conflict of Interest statements are accurate.
	Please confirm you have reviewed this proof to your satisfaction and understand this is your final opportunity for review prior to publication.
AQ: 1	Please insert city and country name for the affiliation "Edgeryders, EE".
AQ: 2	Please insert page number for the quoted text given in the sentence "Following Rheingold (2000) . . . ".
AQ: 3	Please insert expansion for QDA and also insert version of the software if appropriate.
AQ: 4	The term "participant-observation" has been changed to "participant observation" throughout the article as per the usage in the dictionary. Please check and approve.
AQ: 5	Per style, gender bias is to be avoided, please clarify whether "she" could be changed to "he or she" in the sentence "We choose its . . . ".

AQ: 6	Please insert version for the Discourse software if appropriate.
AQ: 7	Please insert version for the computer-assisted qualitative data analysis software if appropriate.
AQ: 8	Please insert expansion for SSNAs.
AQ: 9	Please check the usage of clustering coefficient "*C*" in the sentence "The main component . . . ".
AQ: 10	Per style, gender bias is to be avoided, please clarify whether "she" could be changed to "he or she" in the sentence "She can experiment . . . ".
AQ: 11	Per style, gender bias is to be avoided, please clarify whether "her" could be changed to "his or her" in the sentence "High co-occurrence edges . . . ".
AQ: 12	Please check the usage of the term "infantalized" in the sentence "Refugees wanted the . . . ".
AQ: 13	Per style, gender bias is to be avoided, please clarify whether "she" could be changed to "he or she" in the sentence "She can also . . . ".
AQ: 14	Please confirm whether *Z* denotes informant, if so please suggest whether Z could be changed to informants in the sentence "*Z* are aware . . . ".
AQ: 15	Per style, gender bias is to be avoided, please clarify whether "she" and "her" could be changed to "he or she" and "him or her," respectively, in the sentence "She can add . . . ".
AQ: 16	Please replace section 5.2 with appropriate text in the sentence "Develop the idea . . . ".

Article

Field Methods
1-17

DOI: 10.1177/1525822X20908236
journals.sagepub.com/home/fmx

Semantic Social Networks: A Mixed Methods Approach to Digital Ethnography

Alberto Cottica[1,2], Amelia Hassoun[3], Marco Manca[4], Jason Vallet[5], and Guy Melançon[5] [AQ1]

Abstract

We propose a mixed methods approach to digital ethnographic research. Treating online conversational environments as communities that ethnographers engage with as in traditional fieldwork, we represent those conversations and the codes made by researchers thereon in network form. We call these networks semantic social networks (SSNs), as they incorporate information on social interaction and their meaning as perceived by informants as a group and use methods from network science to visualize these ethnographic data. We present an application of this method to a large online conversation about community provision of health and social care and discuss its potential for mobilizing collective intelligence.

[1] Edgeryders, EE [AQ1: Tallinn, Estonia]
[2] University of Alicante, Campus de San Vicente, Spain
[3] University of Oxford, United Kingdom
[4] SCImPULSE Foundation, CERN openlab, CERN, Geneva, Switzerland
[5] CNRS UMR 5800 LaBRI, University of Bordeaux, France

Corresponding Author:
Amelia Hassoun, University of Oxford, 1 St. Giles, OX1 3JS Oxford, United Kingdom.
Email: amelia.hassoun@oii.ox.ac.uk

Treating conversation platforms as communities in which humans engage in communication and meaning-making (Rheingold 2000) [illegible] ethnography which generates codes that can be analyzed in [illegible]. We define a special type of network representation of ethnographic data, SSNs, and argue that a methodology based on them is accountable to ethnography as a discipline. Its steps, save for the final quantitative analysis layer, carry naturally over from traditional field research to the digital domain. So does ethnography's focus on human communities.

change to: that

SSNs can help discover collective worldviews, address open-ended questions, and scale reasonably well. They show promise as tools to harness collective intelligence, the production, and processing of meaningful information by connected human groups (Lèvy 1997).

We first situate our contribution in digital ethnography and network science literature (see Related Work section). Then, we introduce a data model for SSNs (see A Data Model for Digital Ethnography section). Next, we present data in SSN form from a study on community-provided health and social care services (see An Application: The OpenCare Data subsection). We then illustrate how we used SSNs to aggregate and navigate a large corpus of ethnographic data (see Results and Discussion section). Finally, we reflect on their potential and possible extensions (see Conclusions and Future Improvements section).

Related Work

Digital Ethnographic Methods

Responding to calls for multimodal and network ethnography integrating anthropological participant observation with digital research methods, our ethnographic method maps online conversations to analyze community meaning-making practices (Dicks et al. 2006; Howard 2002; Murthy 2008). Following Rheingold (2000), we define online communities as "social aggregations that emerge from the Net when enough people carry on those public discussions long enough, with sufficient human feeling, to form webs of personal relationships."[AQ2] SSNs encode these webs of relationships. [AQ2: page xx (so citation is (2000:xx)]

We treat online conversational environments (such as digital fora) as communities (Beaulieu 2004; Miller and Slater 2000). We visualize informant contributions in a semantic network, like approaches outlined by Bernard et al. (2016), but with significant differences: Codes are generated through direct ethnographic coding on an online platform and automatically

[AQ3: Qualitative Data Analysis (QDA)]

visualized in network form, in contrast to manual construction of the social network (Howard 2002). The semantic network is therefore not generated from key words but from sustained ethnographic engagement with user contributions. This approach integrates benefits from QDA software[AQ3] (Bernard et al. 2016), iterative meaning-mapping approaches (Dressler et al. 2005), anthropological participant observation, and social network analysis to display a map expressing what community members are talking about and who is talking to whom about what concepts (as interpreted by the researcher). It provides a rich visualization, derived from grounded theory (Bernard 2011; Bernard et al. 2016), to aid in anthropological theorization.

Anthropologists have suggested that network images enhance qualitative understandings of social processes (Burrell 2009; Hannerz 1992; Strathern 1996). Several mixed methods approaches combine qualitative research with social network mapping. Snodgrass et al. (2017) utilize psychometric measurement techniques and interviews to generate measurable qualitative data sensitive to socioculturally specific meaning. Dengah et al. (2018) use survey data and "egocentric social network interviews" to theorize the relationship between social support and online gaming involvement. We employ online ethnographic participant observation[AQ4] rather than offline interviews and construct social networks from observed, rather than reported, online interaction.

[AQ4: Approved.]

Dressler et al. (2005) share our aim: retaining sensitivity to informants' contextual, interactional, and socioculturally specific understandings of concepts while also introducing systematic means of visualizing (and measuring) them. We, however, do not attempt to measure a prehypothesized concept like cultural consonance. Instead, we visualize ethnographic research reflecting the contributions of community members, through codes, in a social network and compute indicators on that network.

SSNs

The notions of semantic network and SSN have been used in different contexts to denote different concepts. Sowa (1983) introduced conceptual graphs as encoding the logical structure of statements also called "semantic networks" (Sowa 1992). His theory focuses on mathematical patterns and rules, and transformations operated on the graph structure emulating human reasoning (Sowa 1983, 2000), mathematically representing knowledge in a manner suitable for deduction (Shapiro 1977; Woods 1975).

Statistics-based approaches known as text mining competed with Sowa's, focusing on co-occurrence patterns of words in documents. Other computational approaches (like unsupervised learning) suggest how groups of terms reflect the semantic structure of documents, often representing documents as weight vectors (Salton and Buckley 1988; Salton et al. 1975). Graphs are utilized when using document vectors to compute similarities between documents (fragments of conversations). Nearest neighbor approaches (Fukunaga and Narendra 1975; Kramer 2013) are often used to induce links between similar documents. Structural characteristics of the graph are interpreted at the corpus level. Jiang et al. (2016) exemplifies the use of these statistical apparatuses to address problems involving large document collections.

We depart from this largely nonethnographic literature by constructing semantic data based on ethnographic coding of community members' conversational themes, elucidated through on-platform int participant-observer researchers. Unlike researchers like R (2010), we do not use natural language processing techniques to theorize how humans co-construct meaning through soc rather than how individual behavior produces network str concepts in more static or single-authored documents like news media relate to each other.

repace hyphen with n-dash [[participant–observer]]

Social networks are traditionally inferred from relati between people (Borgatti et al. 2009), represented as grap are embodied as links between nodes (persons; Wass 1994). Structural metrics (degree, distances, centralities, matrix eigenvalues, etc.) can be interpreted as socially meaningful (Burt 2000; Freeman 1997; Scott 2000).

syllables are: ei-gen-val-ues

We represent ethnographic data as a network. We choose its form so that it encodes information on both what each informant is talking about (as interpreted by the ethnographer) and who she[AQ5] is talking to. So, the network is a semantic and social one—an SSN. It stores the data in a structured form without compromising their rich, contextual character.

SSNs are for human consumption and therefore underpinned by a simple, intuitive ontology. Our representation has only two types of nodes, participants in the conversation and ethnographic codes, and three types of edges: comments-the-content-of (connects a participant to another), authored-posts-annotated-with (connects a participant to a code), and co-occurs-with (connects a code to another). By contrast, graph databases (proposed as early as the 1970s) can and do encode many types of relationships (Shapiro 1977).

[AQ5: Change to she or he is acceptable, but we prefer the female identifier first due to historical gender inequality. The gender neutral/non-binary they' would be even more ideal. Ultimately we leave the final decision to the editors.]

A Data Model for Digital Ethnography

Ethics

OpenCare's conversation was hosted on an instance of the forum software Discourse [AQ6](Atwood et al. 2013–2019), where participants shared care experiences, discussed in-person events, and exchanged best practices. Participants passed through a "consent funnel" before posting, answering required questions to ensure they understood the project and risks associated. The forum conformed to the European Union's General Data Protection Regulation.

We committed to radical methodological openness, discussing research and coding decisions on-platform with informants (Beaulieu 2004; Mortensen and Walker 2002).

Ethnographic Coding: Methods and Implications

We employed iterative, inductive, and reflexive coding processes (Auerbach and Silverstein 2003; Bernard 2011; Saldaña 2015; Sunstein and Chiseri-Strater 2011). An anthropologist (Hassoun) coded 28 hours a month over a one-year period. She performed three coding passes on the data set, with continuous updating of codes based on a detailed codebook with memos and analysis notes. She used both descriptive and in vivo codes.

Hassoun coded directly on-platform via an application we added to the platform's code base. This process created a database of codes and automatically visualized it in network form.

We find ethnographic coding preferable to word count or key words because coding is not just a process of labeling but a clarifying, analytical process in which the ethnographer names and maps the concepts invoked by the community (Saldaña 2015). Informants use different words to mean similar concepts, making a word count less effective. The analytical process undertaken by the ethnographer produces codes that more richly reflect the community's collective intelligence, encoding "summative, salient, essence-capturing, and/or evocative" attributes (Saldaña 2015).

We derive codes from ethnographic participant observation of community members' long-term engagements with each other, analyzing meaning on different levels of analysis. One level is the subjects that community members consider most important and how they relate to each other (visible in the entire network of codes; see Figure 1). Another is a more granular focus on specific topics (manifest in a single code or subnetwork of codes). Yet another level is the focus on a selected group of community members

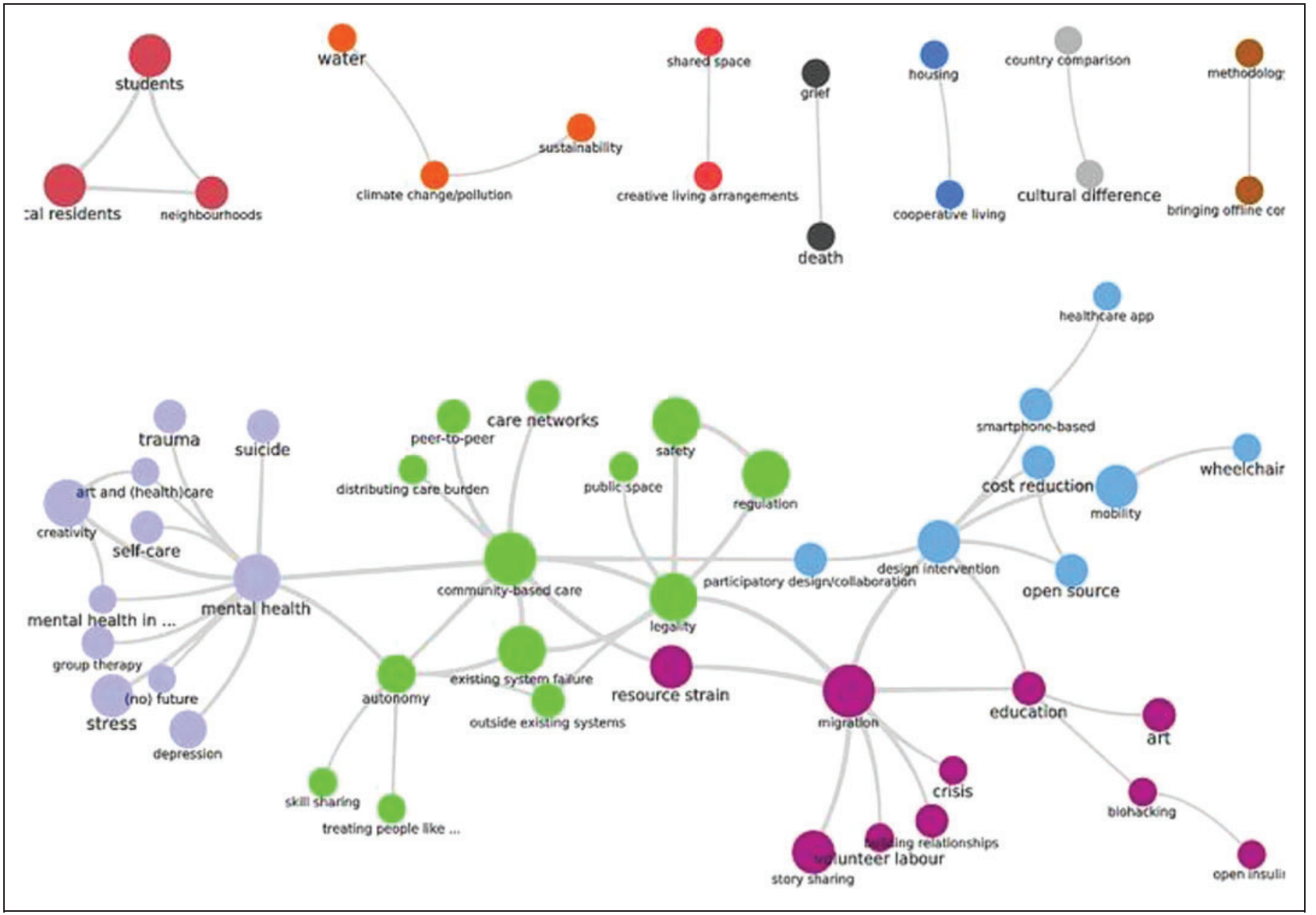

Figure 1. The OpenCare code co-occurrence network (filtered for $k \geq 6$).

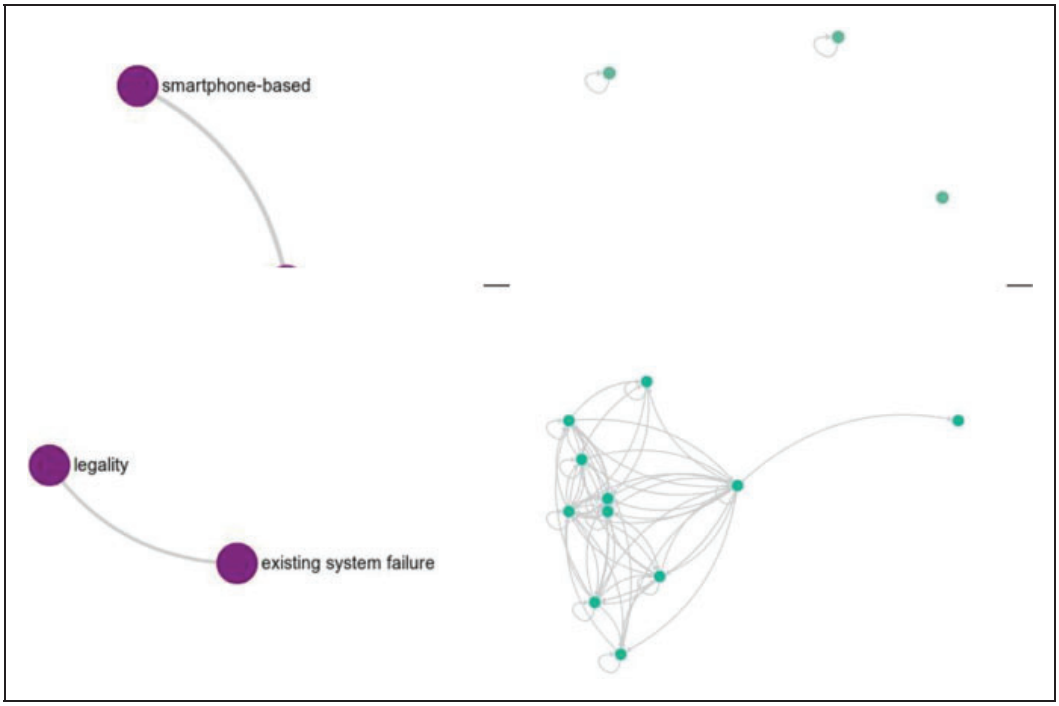

Figure 2. Two edges in the OpenCare semantic network (left) and their associated social networks (right). Structural differences in the latter hint to different degrees of convergence in how the online conversation interprets the former.

(displayed as the network of codes discussed by a selected group; see Figure 2). As contributions and codes are time-stamped, the network of ethnographic codes also allows researchers to see what topics become interesting over time, to whom, and through what kinds of interactions.

Hassoun did not artificially limit the amount of codes assigned (in part so the codes and corpus could be expanded on in the future), instead mapping codes as accurately as possible onto informants' categories of analysis. Because we employ grounded theory and had community members participating from a wide range of cultural and institutional contexts, we did not predefine any codes based on existing theories.

On-platform coding produces a digital codebook, making coding decisions transparent and codes easily editable. Visualizing the semantic network enables iterative coding processes, illuminating which codes might be redundant or need forking. Both elements enable multipl delete comma work on the same large corpus in a coherent way, with unique identifiers separating each researcher's codes. This makes our method partially scalable. It offers a clear benefit over computer-assisted qualitative data analysis software[AQ7], which is closed, mostly proprietary, and notoriously difficult for multiple researchers to use.

These methods also make the ethnographer's interpretive process more visible. Coding is a process of reflexive interpretation that requires moving between the positionality of the researcher and the worldviews of informants (Rosaldo 1992). Keeping a detailed, open codebook and memos makes this more transparent than in traditional ethnographic studies. These coding practices and visualizations utilize ethnography's ability to tease out collective beliefs and practices while rendering the researcher's situatedness and partiality visible. Future studies employing multiple ethnographers will enable comparison of SSNAs[AQ8] generated by different researchers, further shedding light on this interpretive process.

[AQ8: Change to 'SSNs' and expansion is Semantic Social Networks]

Contributions

SSN-based ethnographies start with posts/comments on the social networking platform. We call contribution a testimony in written form (interview transcript, post on an online forum, etc.). A minimum viable structure for encoding a contribution as primary data includes:

Contribution ID: The contribution's unique identifier.
Text: The contribution's complete text.
Author ID: A unique identifier for the informant that contributed the text.
Target ID: A unique identifier for the informant that the text is addressed to.
Date and time.
Annotations.

Ethnographers associate snippets of texts in contributions to key words called codes. This generates an ontology representative of the corpus. We call annotation the atomic result of this activity. A minimum viable structure for annotations includes:

Annotation ID: The annotation's unique identifier.
Contribution ID: The unique identifier of the post or comment that this annotation refers to.
Snippet: The part of the text in the contribution that the researcher wishes to associate with the code.
Code: The ethnographic code associated with the snippet.
Author ID: Unique identifier for the researcher that produced the annotation.
Date and time.

This representation induces a network where the nodes are informants and edges represent interactions. Codes—associated with the interaction via annotations—encode the semantics of that interaction. We call this an SSN. We propose it is general enough to fit evidence from most ethnographies, while structured enough to be encoded into a data set.

An Application: The OpenCare Data

The OpenCare project explores how communities provide health and social care when neither states nor businesses can or will serve them. We began with the research question: What do people do when existing health and social care systems no longer provide care? Data were gathered from an online forum where individuals discuss their care experiences. We used the method described in A Data Model for Digital Ethnography section to code them and build the corresponding SSN. We then built a social and a semantic network from the coded data.

The OpenCare Social Network

Online conversation induces a social network where nodes are community members and edges encode interaction. For users A and B, we induce a connection $A \rightarrow B$ if A has commented B's content at least once. This network is directed ($A \rightarrow B \neq B \rightarrow A$) and weighted (the edge $A \rightarrow B$ has a weight of k if A has commented B's content k times). The OpenCare corpus has 332 nodes and 1,265 edges.

The main feature of this network is a clear core–periphery structure. Almost all participants are connected to the giant component, so information can flow freely across the network. The giant component itself is not obviously resolved into distinct subcommunities (its modularity value is 0.38 Newman and Girvan [2004]). These structural features allow us to infer that most opinions expressed in the forum have been expressed in a public space that everybody participates in. There are no signs of isolation of individuals, nor of balkanization of the conversation.

SSNs can also be represented in ways that emphasize the semantics of the online conversation. The representation that proved most useful to ethnographic research is what we call the co-occurrence network. Its nodes are codes. Whenever two codes occur in annotations that refer to the same post, they are said to co-occur in the same post, and an edge is induced between them. This network is undirected ($A \rightarrow B \equiv B \rightarrow A$) and weighted (the edge has a weight of k if A co-occurs with B on k different posts or

comments). We can think of the co-occurrence network as an association map between the concepts expressed by the codes. A higher edge weight k indicates a stronger group-level association between the two codes connected by the edge.

The annotations on the OpenCare corpus induce a co-occurrence network with 1,248 nodes and 16,727 edges. The main component is formed of 1,234 nodes and 16,702 edges and shows a small-world structure (Watts and Strogatz 1998) with a high average clustering coefficient $C^{*} = 0.696$.[AQ9]

Results and Discussion

Filtering the Co-occurrence Network for a High-level View

Rather than representing the point of view of an individual, the co-occurrence network encodes contributions from informants as a group in conversation as interpreted by an ethnographer. The resulting concept map, therefore, does not simply aggregate the association patterns of individuals like a survey; it is the product of the interaction across participants. Edge weight k, then, represents the strength with which the conversation associates the codes connected by that edge.

Filtering the graph by higher value of k allows the researcher to see the strongest associations between codes made by informants as a group. She can experiment with different values of k, starting from a low value and increasing until the graph simplifies enough to be interpretable[AQ10]. For the OpenCare data set, filtering edges by $k \geq 6$ yields a co-occurrence network with 60 codes and 72 edges, which lends itself well to visual inspection (Figure 1). [AQ10 : Same answer as AQ5]

From it, one can see the structure of community members' concerns. Consider the cluster with *legality* (in green): We find *existing system failure* and *regulation*, reflecting the preoccupation of some informants that community health-care initiatives and technological innovations outside of existing systems (much needed when systems fail) turn out to be illegal and therefore difficult to implement. We also find *safety*, reflecting the acknowledgment that regulation is often there for a reason.

We can also see isolated but intense conversations visualized as islands in the high co-occurrence network: The network *death*, *grief*, and *online memorials* (visible at a lower co-occurrence level) appear unconnected to the rest of the network, indicating a deeply discussed single issue. In this case, community members intensively discussed this issue on one highly active thread, but the topic was not discussed more widely across the

[AQ11: Same answer as AQ5]

platform. In the next section, we describe how to tell whether a discussion is driven by a small number of community members or a larger group.

A researcher can choose to look at the network of associations around a topic of interest at a more granular level by clicking on the link between two codes to view all community contributions containing both codes (like *design intervention* and *cost reduction*) to see what specific innovations the community has devised.

The method allows for rich analysis on multiple levels, retaining the granularity that makes ethnographic research so powerful. High co-occurrence edges in the semantic network illuminate connections that might be invisible at a smaller scale of analysis, allowing the ethnographer to visualize and understand her informants' concerns and how they relate to each other[AQ11]. Without the co-occurrence network, vital interconnections made by informants would have been missed; without the detailed ethnographic data, the meaning behind those connections would be lost.

A detailed discussion is out of scope of this methods-focused article, but in the OpenCare project, this method led to key insights into informants' beliefs, desires, innovations, and concerns. Centrally, people, facing the collapse of existing health and social care institutions, reach for what we term "collective autonomy": feeling empowered to solve their own problems while in a community-based framework (Hassoun 2017). This finding has clear implications for states and nongovernmental organizations trying to help people in crisis. Care solutions that treat people as helpless or remove them from a community context will likely fail. Refugees wanted the tools to collaboratively build their own temporary living spaces and markets rather than being infantalized[AQ12]; mental health patients found helping others in their community therapeutic. Solutions that connected people with others with compatible skills gave them tools and space to experiment, or strengthened care networks in communities were most useful to people seeking care outside of existing health and social care systems.

Some of the network's properties have straightforward interpretations and can be used to validate or extend the researcher's conclusion. A researcher can use edge weight to get a precise idea of how strong the association between any two codes is. She can also use community detection algorithms to get a quantitative indicator of how neatly a problem resolves into subissues[AQ13]. We applied to the network in Figure 1 the Louvain community detection algorithm (Blondel et al. 2008): It is highly modular (with a modularity of 0.64) and presents clearly distinguishable communities of codes identified by color.

[AQ12: Infantalized is the correct term. For clarity, we could add: "rather than being infantalised by states or non-governmental organizations"]

[AQ13: Same answer as AQ5]

Enriching Semantic Information with Social Network Structure Information

The OpenCare social and semantic networks, as described in The OpenCare Social Network and SSNs subsections, are interlinked by the data structure defined in A Data Model for Digital Ethnography. This enriches semantic information with information on the structure of the social network. For example, we can check that the social network underpinning any one edge in the semantic network is connected. A connected social network signals that informants who have made the connection between those two codes are in conversation with each other: *Z* [AQ14]are aware of each other's existence and have had the opportunity to interact around that particular connection to arrive at an interpretation of its nature and importance for the problem at hand. A disconnected one signals that they never conversed at all: They agree the two codes are connected but might have different interpretations of that connection unimpacted by each other's views. In Figure 2, there are four informants who have mentioned both smartphone-based and health-care app (six co-occurrences), and they are not interacting directly with each other. There are 11 who have mentioned both legality and existing system failure (nine co-occurrences), and they are all connected in a dense network of direct interaction. The latter association has the potential for being supported by a consensus resulting from the conversation, not unlike what happens in Wikipedia (Laniado et al. 2011), whereas the former does not.

[AQ14: Replace with 'informants']

Conclusions and Future Improvements

SSNs show promise as a digital social science research method aimed at capturing collective intelligence and making ethnography a more collaborative discipline. They deal well with open questions and novelty (like traditional ethnography) and handle hundreds of informants (like quantitative surveys). When combined with open standards and open data, they could perhaps attempt to handle thousands of informants.

SSNs pave the way for replication, reuse, and extension of ethnographic studies as well as larger-scale studies. An ethnographer can pull a colleague's annotations and codebook, increasing the clarity and accountability of the research process. She can add her own coded corpus and use the combination of annotated corpora to produce a new study[AQ15]. Accurate documentation of the code ontology allows ethnographers to work on projects that would be too large for a single ethnographer to tackle. Finally,

[AQ15 : Same answer as AQ5]

SSNs help enable longitudinal online ethnography, as an online conversation could be revamped yearly to keep track of how its collective point of view evolves.

These practices require a cultural shift from practitioners. Ethnographers tend to work alone and seldom disclose access to coded interviews and field notes. The process of coding in ethnography follows standards that are project specific and often not made public. There are few naming conventions for codes followed by all ethnographers, few codebooks published in electronic form, no accepted speci d so on. We suggest that ethnographers embrac d publishing open data. Open data are data that are (1) machine-readable, (2) published under licenses that allow their reuse, and (3) documented with appropriate metadata.[1]

change commas to semicolons

The payoff of such a shift is substantial. We could imagine a version of Eurobarometer based on an open online conversation. Instead of answering multiple-choice questions, vulnerable to framing biases (Tversky and Kahneman 1981), informants would discuss their perception of Europe, allowing researchers to discover novel patterns of association and detect the fading of old ones.

Our method could be further improved along the following lines:

[AQ16: Enriching Semantic Information with Social Network Structure Information]

1. Develop the idea, introduced in section 5.2 [AQ16], of applying existing social theory on the social network topology to derive "interest scores" on individual informants and connections in the semantic network.
2. Apply alternative ways to measure edge (association) strength k in the co-occurrence network. For example, $k(A \rightarrow B)$ could encode the number of informants who have authored contributions coded with both codes A and B or the number of separate threads that contain at least one contribution with it. Different measures of edge strength have different interpretations, so they allow different perspectives on the data corpus.
3. Observe and model the online conversation as a dynamic system. Stochastic actor-oriented models might be a good place to start despite known limitations (Snijders 1996).
4. Multilingual studies: The code ontology can be structured as a hierarchy so that codes with the same meaning in different languages are entered in the secondary data as children of the same parent code. The code co-occurrence network would be drawn between parent nodes, allowing both an all-languages view on data and across-

languages comparisons. This work would allow for analysis of generalizability of codes across languages and potentially different cultural understandings of concepts, although cultural difference would need to be explicitly theorized as an object of analysis.

5. Ethnographic coding as compression: Empirically, in OpenCare, the rate at which new codes are produced declined, approaching zero. This suggests that, after a certain number of iterations, the online conversation produced an ontology (nearly) sufficient to code all future contributions. This phenomenon is reminiscent of file compression in computer science. If the analogy holds, a dynamic analysis of the conversation could identify moments of significant evolution by spikes in the growth rate of the codes dictionary as new ideas are suddenly generated. We could estimate the overall relatedness and novelty of conversations. We intend to refine and test this hypothesis on more corpora of similar size and scope.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The OpenCare project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 688670).

Supplemental Material

Supplemental material for this article is available online.

Note

1. We have released the OpenCare data set as open data (https://doi.org/10.5281/zenodo.164970; https://github.com/opencarecc/opencare-data-documentation).

References

Atwood, J., R. Ward, S. Saffron, and N. Lalonde. 2013–2019. Discourse: Civilized discourse construction kit. https://www.discourse.org/ (accessed July 28, 2019).

Auerbach, C., and L. B. Silverstein. 2003. *Qualitative data: An introduction to coding and analysis*. New York: NYU Press.

Beaulieu, A. 2004. Mediating ethnography: Objectivity and the making of ethnographies of the Internet. *Social Epistemology* 18:139–63.

Bernard, H. R. 2011. *Research methods in anthropology: Qualitative and quantitative approaches*, 5th ed. Walnut Creek, CA: AltaMira.

Bernard, H. R., A. Y. Wutich, and G. W. Ryan. 2016. *Analyzing qualitative data: Systematic approaches*, 2nd ed. Thousand Oaks, CA: Sage.

Blondel, V. D., J. L. Guillaume, R. Lambiotte, and E. Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008:10008.

Borgatti, S. P., A. Mehra, D. Brass, and G. Labianca. 2009. Network analysis in the social sciences. *Science* 323:892–95.

Burrell, J. 2009. The field site as a network: A strategy for locating ethnographic research. *Field Methods* 2:181–99.

Burt, R. S. 2000. The network structure of social capital. *Research in Organizational Behavior* 22:345–423.

Dengah, H., J. Franois, J. Snodgrass, R. Else, and E. Polzer. 2018. The social networks and distinctive experiences of intensively involved online gamers: A novel mixed methods approach. *Computers in Human Behavior* 80:229–42.

Dicks, B., B. Soyinka, and A. Coffey. 2006. Multimodal ethnography. *Qualitative Research* 6:77–96.

Dressler, W. W., C. D. Borges, M. C. Balierio, and J. E. dos Santos. 2005. Measuring cultural consonance: Examples with special reference to measurement theory in anthropology. *Field Methods* 17:331–55.

Freeman, L. C. 1997. A set of measures of centrality based on betweenness. *Sociometry* 40:35–41.

Fukunaga, K., and P. M. Narendra. 1975. A branch and bound algorithm for computing k-nearest neighbors. *IEEE Transactions on Computers* 100:750–53.

Hannerz, U. 1992. *The global ecumene as a network of networks*. New York: Routledge.

Hassoun, A. 2017. Collective autonomy. An ethnography of open care services. https://doi.org/10.5281/zenodo.1186606 (accessed March 25, 2018).

Howard, P. 2002. Network ethnography and the hypermedia organization: New media, new organization, new methods. *New Media and Society* 2:517–35.

Jiang, K., G. A. Barnett, and L. D. Taylor. 2016. News framing in an international context: A semantic network analysis. *International Journal of Communication* 10:3710–36.

Kramer, O. 2013. *Dimensionality reduction with unsupervised nearest neighbors*. Berlin, Germany: Springer.

Laniado, D., T. Tasso, Y. Volkovich, and A. Kaltenbrunner. 2011. When the Wikipedians talk: Network and tree structure of Wikipedia discussion pages. Fifth

International Conference on Weblogs and Social Media (ICWSM). https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2764/3301 (accessed August 18, 2017).

Lèvy, P. 1997. *Collective intelligence: Mankind's emerging world in cyberspace*. New York: Perseus books.

Miller, D., and D. Slater. 2000. *The Internet: An ethnographic a* change to: London: UK: Bloomsbury Press.

Mortensen, T., and J. Walker. 2002. *Blogging thoughts: Personal publication as an online research tool*. Oslo, Norway: University of Oslo change to Oslo:

Murthy, D. 2008. Digital ethnography: An examination of ogies for social research. *Sociology* 42:837–55.

Newman, M. E. J., and M. Girvan. 2004. Finding and evaluating community structure in networks. *Physical Review E* 69:1–16.

Rheingold, H. 2000. *The virtual community: Homesteading on the electronic frontier*. Cambridge, MA: MIT Press.

Rosaldo, R. 1992. *Culture and truth: Remaking* change to: Boston: MA: Beacon Press.

Roth, C., and J. Cointet. 2010. Social and semantic coevolution in knowledge networks. *Social Networks* 32:16–29.

Saldaña, J. 2015. *The coding manual for qualitative researchers*, 3rd ed. Newbury Park, CA: Sage.

Salton, G., and C. Buckley. 1988. Term-weighting approaches in automatic text retrieval. *Information Processing & Management* 24:513–23.

Salton, G., A. Wong, and C.-S. Yang. 1975. A vector space model for automatic indexing. *Communications of the ACM* 1:613–20.

Scott, J. P. 2000. *Social network analysis: A handbook*, 2nd ed. Newbury Park, CA: Sage.

Shapiro, S. C. 1977. Representing and locating deduction rules in a semantic network. *ACM SIGART Bulletin* 63:14–18.

Snijders, T. A. B. 1996. Stochastic actor-oriented models for network change. *Journal of Mathematical Sociology* 21:149–72.

Snodgrass, J., H. J. F. Dengah, M. Lacy, A. Bagwell, M. Van Oostenburg, and D. Lende. 2017. Online gaming involvement and its positive and negative consequences: A cognitive anthropological cultural consensus approach to psychiatric measurement and assessment. *Computers in Human Behavior* 66: 291–302.

Sowa, J. F. 1983. *Conceptual structures: Information processing in mind and machine*. Reading, MA: Addison-Wesley.

Sowa, J. F. 1992. Conceptual graphs as a universal knowledge representation. *Computers & Mathematics with Applications* 23:75–93.

Sowa, J. F. 2000. *Knowledge representation: Logical, philosophical, and computational foundations, vol. 13*. Pacific Grove, CA: Brooks/Cole.

Strathern, M. 1996. Cutting the network. *The Journal of the Royal Anthropological Institute* 2:517–35.

Sunstein, B. S., and E. Chiseri-Strater. 2011. *FieldWorking*, 4th ed. Bedford, MA: St. Martins.

Tversky, A., and D. Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211:453.

Wasserman, S., and K. Faust. 1994. *Social network analysis, methods and applications. Structural analysis in the social sciences*. Cambridge, MA: Cambridge University Press.

Watts, D. J., and S. H. Strogatz. 1998. Collective dynamics of "small-world" networks. *Nature* 393:440–42.

Woods, W. A. 1975. What's in a link: Foundations for semantic networks. In *Representation and understanding: Studies in cognitive science*, ed. W. A. Woods, 35–82. Cambridge, MA: Elsevier.