Appraising discrepancies and similarities in semantic networks using concept-centered subnetworks

Darkhan Medeuov[1,3]* , Camille Roth[2], Kseniia Puzyreva[3] and Nikita Basov[3]

*Correspondence:
darkhan.medeuov@nu.edu.kz
[3] Centre for German
and European Studies, St.
Petersburg State University,
St. Petersburg, Russia
Full list of author information
is available at the end of the
article

Abstract

This article proposes an approach to compare semantic networks using *concept-centered* sub-networks. A concept-centered sub-network is defined as an induced network whose vertex set consists of the given concept (ego) and all its adjacent concepts (alters) and whose link set consists of all the links between the ego and alters (including alter-alter links). By looking at the vertex and link overlap indices of concept-centered networks we infer semantic similarity of the underlying concepts. We cross-evaluate the semantic similarity by close-reading textual contexts from which networks are derived. We illustrate the approach on written and interview texts from an ethnographic study of flood management practice in England.

Keywords: Semantic networks, Flood management, Computational text analysis

Introduction

Text can be represented as a semantic network, with vertices being words and links being counts of word co-occurrences within a given distance between words. Research in cultural sociology often views semantic networks as a model of knowledge/culture underpinning the production of text (Carley 1986; Carley and Newell 1994; Lee and Marin 2015). This allows researchers to infer knowledge structure using the tools of network analysis (Abbott et al. 2015; Roth and Cointet 2010). Currently, an issue of particular interest is how different knowledge/culture systems (for instance institutional-field and local knowledge, see Basov et al. 2019) interplay with each other. The present paper proposes an approach to examine this interplay at the meso-level of particular concepts (word lemmas) which gain different meanings across different knowledge/culture systems.

We draw on a new text dataset on professional and local flood management knowledge collected during one of this paper's authors ethnographic fieldwork in England. Flood management in England as a knowledge domain provides an apt example because, until recently, it exclusively relied on institutionalized professional knowledge. In recent decades, however, flood management has become more sensitive to "local knowledges" and started seeking ways to engage local actors (McEwen and Jones 2012; Nye et al. 2011;

Wehn et al. 2015). Becoming stakeholders in flood risk management, local communities/activists are expected to adhere to professional knowledge and language. They, however, rarely take professional knowledge at "face value", but rather creatively reuse it to fit local context (Nye et al. 2011; Wehn et al. 2015). Our data comes from one such flood-prone area where flood management agencies and local community groups collaborate to manage flood risks. To examine which professional concepts are used by local actors, we represent professional and local knowledge as separate semantic networks and then compare sub-networks centered on the concepts that are shared across professional and local semantic networks.

The paper is organized as follows. In the next section, we introduce the idea of concept-centered sub-networks and lay out reasons why researchers might better understand semantic networks focusing on their concept-centered components. Then, we describe our two-step analytical approach, first by showing how to find potentially interesting concept-centered sub-networks and then how to highlight their similarities and differences. We illustrate the approach using English flood management data. We conclude by discussing the main results and outlining future work prospects.

Semantic and concept-centered networks

A semantic network, like any other network, is defined as a couple of sets $G = (V, E)$. V lists all vertices, and $E \in V \times V$ lists all connected pairs of vertices. A network can also be defined by an adjacency matrix M of dimension $|V| \times |V|$ where entry m_{ij} indicates the presence of a link between vertices i and j. In the most simple case M is a binary and symmetric matrix: $m_{ji} = m_{ij} \in \{0, 1\}$. Semantic networks, however, can include co-occurrence counts, in which case entries of their adjacency matrices are not binary but positive integers, $m_{ij} \in \mathbb{N}$. Researchers nonetheless often binarize such co-occurrence matrices by using some threshold value, τ, $(m_{ij} \geq \tau \rightarrow m_{ij} = 1) \wedge (m_{ij} < \tau \rightarrow m_{ij} = 0)$ (Dianati 2016; Cantwell et al. 2020).

The typical case we address here consists in comparing semantic networks associated with professional vs. local knowledge. There exist many ways to compare networks, however, we argue that semantic networks specifically may be better understood by comparing different substructures they consist of (e.g. network motifs) (Choobdar et al. 2012; Pržulj 2007; Milo et al. 2002). We further argue that comparison of knowledge systems can be narrowed down to the use of particular concepts, which at the level of semantic networks corresponds to concept-centered sub-networks. For example, knowing that the concept "*flood*" is linked to different concepts in professional and local semantic networks[1] may hint at the fact that *flood* has different meanings in professional and local knowledge systems.

Thus, in practice, we compare concept-centered networks by appraising the similarities of their vertex and link sets. We define a concept-centered network as follows. For a given actor a and given concept c the concept-centered network $C_{a,c} = \{V_{a,c}, E_{a,c}\}$ is an induced network whose vertex set $V_{a,c}$ consists of the given concept (ego) and its

---

[1] That is, the concept tends to co-occur with different other concepts in professional and local texts.

adjacent concepts (alters) and whose link set $E_{a,c}$ consists of all links in E between pairs of vertices in $V_{a,c}$.

We measure similarity using vertex and link overlap indices, which generally indicate how many shared elements two sets have relative to the cardinality of the smaller set. We use this index instead of the more common Jaccard index because in practice local actors draw upon professional knowledge, not the other way around. In other words, we are examining the part of concepts and linkages in professional knowledge that local actors draw on when adjusting professional knowledge to their specific local perspective[2]. We further exclude the ego concept and ego-incident links in the computation of indices because keeping them induces a lower bound. While this lower bound is somewhat negligible for the vertex overlap, it can be substantial for the link overlap. We denote as $V_{a,c}^*$ and $E_{a,c}^*$ the vertex and link sets that exclude ego and ego-incident links. Focusing on the c-centered semantic networks of actors a and a', and denoting the overlap index of two sets X and Y as:
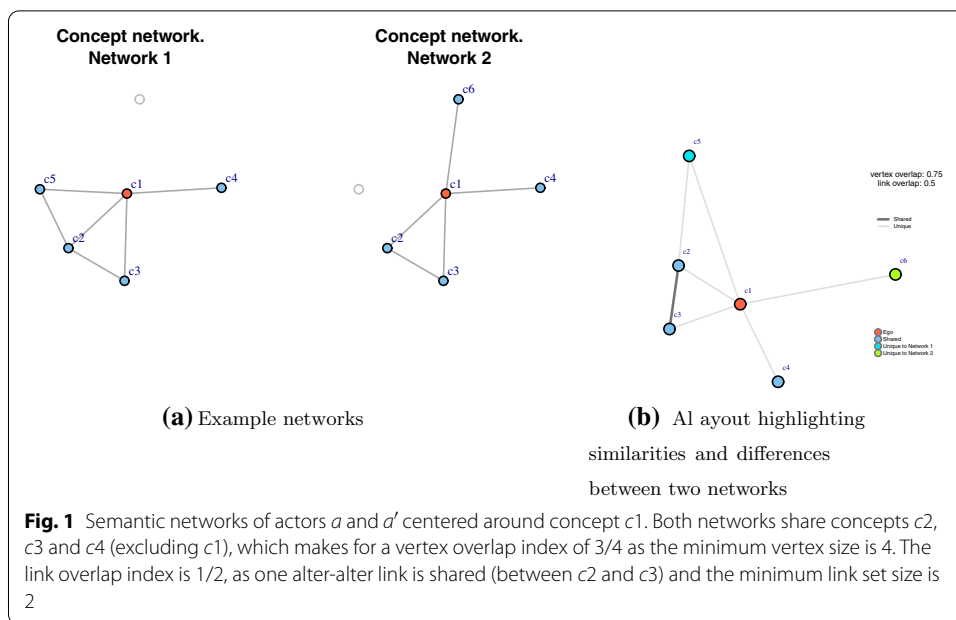
$$\omega(X, Y) = \frac{|X \cap Y|}{\min(1, |X|, |Y|)}$$

we define their vertex overlap index as $\omega(V_{a,c}^*, V_{a',c}^*)$, and their link overlap index as $\omega(E_{a,c}^*, E_{a',c}^*)$. The denominator of the overlap index definition guarantees consistence in case the central concept c has no alter, producing an overlap of 0 by construction (this corresponds to trivial concept connection configurations which we are nonetheless generally not interested in).

Along with the analytical approach we also propose a visualization approach to highlight similarities and discrepancies between two concept-centered networks. We first combine concept-centered sub-networks into a "union" network by gathering all vertices and links. The union graph helps to position vertices and keep these positions in the visualization of each network separately. We also assign a double weight to links shared by both actors: this plays a functional role as the force-directed layout makes vertices incident to shared links more attracted to each other. This not only puts pairs of concepts that are connected for both actors closer to each other, but also spatially separates them from concepts that are not, thus further highlighting commonalities and discrepancies between networks.

Figure 1 illustrates this idea. Suppose we have a pair of actors and their respective concept-centered networks (1a). The ego-concept in both networks is *c1*. The first actor connects *c1* to *c2*, *c3*, *c4*, and *c5*; the second actor does not connect *c1* to *c5*, but to *c6*. Thus, the vertex overlap index between these concept-centered networks is 0.75 (3 shared vertices out of 4 vertices). Besides, there is a difference in the alter-alter links: while both actors connect *c2* and *c3*, only the first actor connects *c2* and *c5*. Therefore, the link overlap similarity between them is 0.5 (1 shared alter-alter link out of 2).

The visual representation we are proposing is presented in Fig. 1b: shared links between concepts tend to clump together near ego because their weight is double, which results in larger "attraction power" of links to these concepts in terms of Fruchterman-Reingold

---

[2] The size of local semantic network is generally smaller than that of professionals, mostly due to different corpus sizes, which, in turn, is due to locals demonstrating less elaborate verbalised knowledge of the subject.

(a) Example networks

(b) Al ayout highlighting similarities and differences between two networks

Fig. 1 Semantic networks of actors a and a' centered around concept $c1$. Both networks share concepts $c2$, $c3$ and $c4$ (excluding $c1$), which makes for a vertex overlap index of 3/4 as the minimum vertex size is 4. The link overlap index is 1/2, as one alter-alter link is shared (between $c2$ and $c3$) and the minimum link set size is 2

layout. Along the same logic, non-shared concepts are such that by construction no links may be shared across actors, thus the corresponding links pointing to such concepts have lower weights—which, in turn, repeals them farther away from ego. Furthermore, concepts connected with shared links appear closer to each other than those connected by non-shared links, again, because uniqueness of the latter links makes their weight lower. In addition, concepts linked with the "shared core", even if they and their links are not shared (like $c5$) are placed closer to ego than "lone-standing" non-shared concepts (like $c6$).

In what follows we apply the two-stage approach described above to compare concepts across semantic networks and discuss some particularly interesting concept-centered comparisons.

Data description

Our data come from an ethnographic study focusing on two local flood management groups located in the County of Shropshire, England. Professional knowledge is represented with a collection of documents issued by flood management agencies and authorities (around 316,000 words in total). Local knowledge is represented with semi-structured interviews with 15 members of the two *local flood groups* (LFG)—local activist groups involved in flood risk management in two villages. We denote these groups as *LFG1* and *LFG2*, respectively. The interviews comprise 186,000 words in total, with the average word number per interview being around 13,000.

Note that we represent professional knowledge with one network (further, "professional network"). This reflects a sort of "universality" of professional knowledge, as we assume that the content of official documents should reflect some general consensus among professionals. For locals, on the other hand, we allow each interviewee to have their own semantic network. We do so because we are interested in looking at how
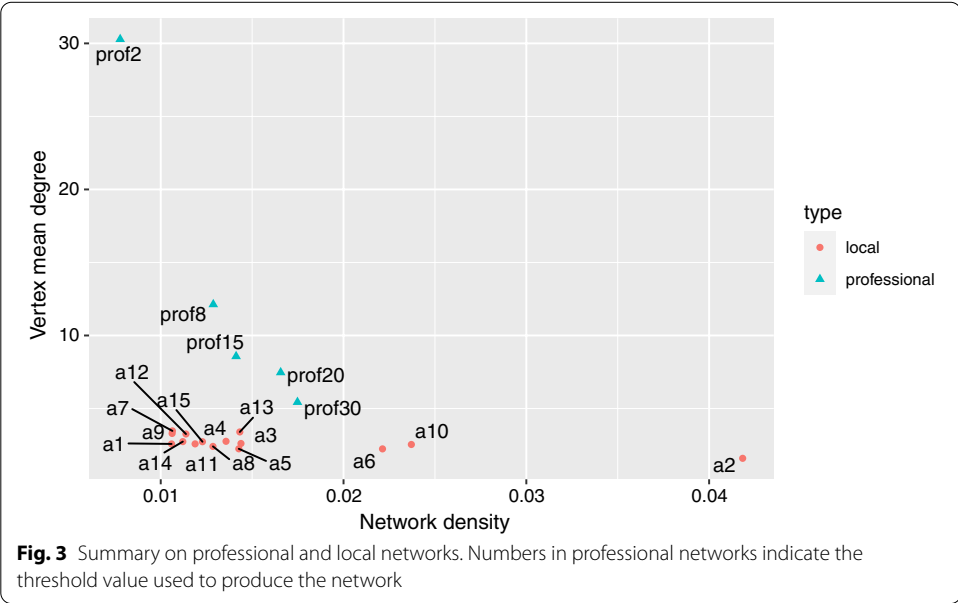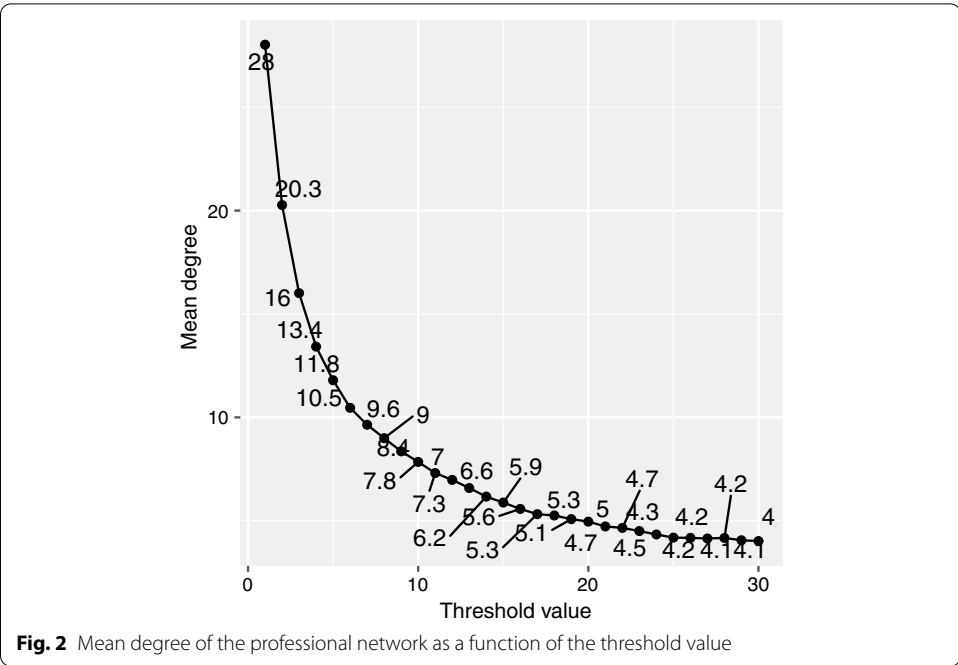
Table 1 Local and professional networks summary statistics

Network	Vertices	Links	Density	Mean degree	Max degree	Most connected term
a1	244	314	0.01	2.57	37	people(n)
a2	39	31	0.04	1.59	5	contact(v)
a3	182	237	0.01	2.6	20	flood(n)
a4	204	281	0.01	2.75	53	water(n)
a5	158	177	0.01	2.24	22	people(n)
a6	102	114	0.02	2.24	13	flood(n)
a7	327	567	0.01	3.47	40	water(n)
a8	188	226	0.01	2.4	19	water(n)
a9	311	512	0.01	3.29	33	flood(n)
a10	108	137	0.02	2.54	25	flood(n)
a11	218	281	0.01	2.58	25	people(n)
a12	287	467	0.01	3.25	31	flood(n)
a13	238	404	0.01	3.39	59	flood(n)
a14	245	335	0.01	2.73	45	flood(n)
a15	223	304	0.01	2.73	45	flood(n)
Professional: 2	3893	58947	0.01	30.28	1627	flood(n)
Professional: 9	943	5718	0.01	12.13	549	flood(n)
Professional: 15	608	2606	0.01	8.57	358	flood(n)
Professional: 20	452	1688	0.02	7.47	288	flood(n)
Professional: 30	312	848	0.02	5.44	203	flood(n)

particular local actors borrow concepts from professional knowledge. We denote local networks with lowercase *a*'s followed by a number (e.g., *a1* or *a2*).

We produce semantic networks from texts as follows. First, raw texts are tokenized and part-of-speech (POS) tagged using the UDpipe package (Wijffels 2019): we convert words into lemmas and combine lemmas with their POS tags to produce unique concept identifiers (e.g., "flood(v)" as the verb and "flood(n)" as the noun), which we refer to as concepts in this paper. Research assistants have manually inspected the corpus checking for machine-missed stopwords (usually numbers e.g., "60s"), transcribed artefacts of oral speech (e.g., "aha", "eh"), set phrases ("bear mind", "couple time"), incorrectly recognized words, informants' real names, and the same words that have different spelling. All such instances have been replaced with either correct versions or a generic placeholder.

We count co-occurrences using the sliding window approach: concepts count as co-occurred every time they appear within 8-concept vicinity from each other, unless separated by a full stop mark. This yields weighted co-occurrence networks. We then filter these networks from all non-nouns, non-verbs, non-adjectives as well as from trivial verbs (e.g., "do" or "make"), leaving only vertices related to adjectives, nouns, and non-trivial verbs. We take this step to (a) reduce the amount of information to process and (b) to focus on informative parts of speech. Finally, we binarize co-occurrence counts using the threshold of 2 for both professional and local networks, that is we link all the pairs of concepts which co-occurred at least 2 times (see next section for a discussion). The total number of links in the resulting professional network is 58,947, while in the local networks it is 300, on average. The mean degree for the professional network is about 30, while the average mean degree for local networks is 2,69. Table 1 summarizes features of local and professional networks.

Fig. 2 Mean degree of the professional network as a function of the threshold value

Fig. 3 Summary on professional and local networks. Numbers in professional networks indicate the threshold value used to produce the network

On the choice of binarization threshold value

Figure 2 shows mean degrees of different professional networks against threshold values used to produce them, while Fig. 3 juxtaposes various professional network specifications with the local networks in terms of their mean degree and density. The threshold value clearly affects topology of the network and, therefore, should have an impact on the chances of professional and local concept-centered networks to overlap. It is therefore important to ground the choice of binarization threshold.
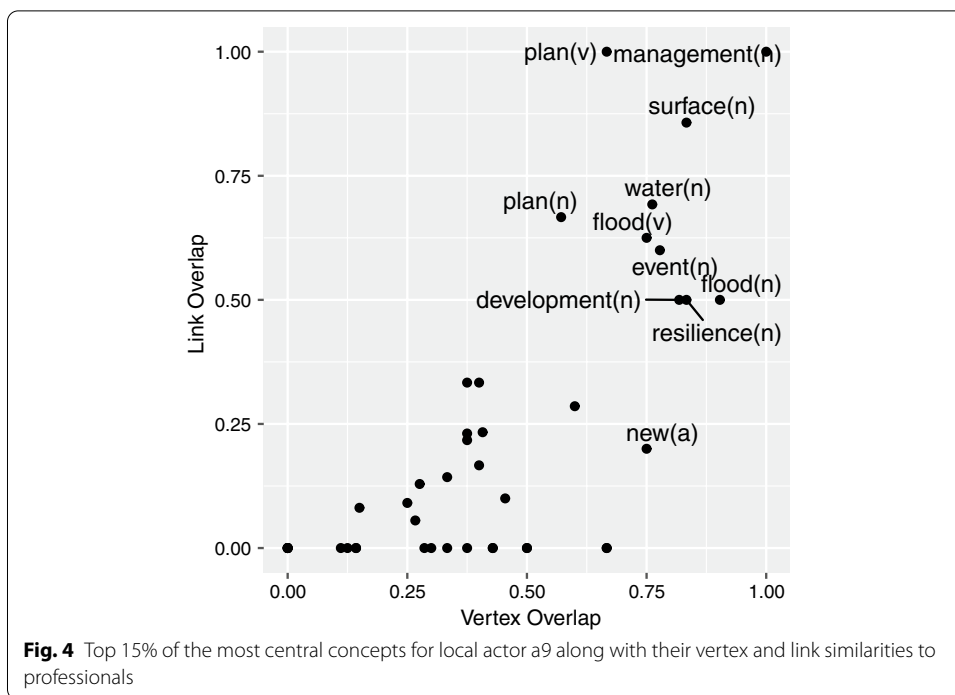
In this paper, we choose to work with the same threshold for both professional and local networks. The reason is twofold: first, it lowers the number of parameters that we as researchers use to process and possibly alter the raw network data. Second, the justification of the threshold depends on the interpretation of co-occurrences in general. On the one hand, higher co-occurrence counts may be induced by the larger size of the professional corpus in comparison with the local corpus. On the other hand, higher co-occurrence counts may be due to a higher density of linkages between concepts in professional vs. local networks because for instance professional knowledge is more elaborated. If we assume that both can be the case, a correctly chosen threshold would be able to both preserve the higher connectivity of professional knowledge networks and reduce the potential bias introduced by using corpora of different sizes to define professional semantic network and various semantic networks of each local actor.

Let us consider a hypothetical situation in which the professional network strictly comprises all local networks, which is actually quite realistic especially in terms of links. In other words, local networks are all sub-networks of the professional network. Two simple scenarios are possible. First scenario, professional corpuses cover a much *broader* set of topics than are covered by locals: the discrepancy in text size simply reflects the fact that many more issues are being addressed by experts and that only a fraction of the expert data deals with issues that are present in the local data. In that case, we argue that the expert network centered around discourses mentioning a given issue should be directly comparable with the local network centered around that same issue. Second scenario, professional corpuses cover the same set of topics but in much *denser* manner: the discrepancy in text size reflects the utterance by professional of more sentences around the same topics, and thus a richer set of connections for these same topics. In that case, the same threshold should also apply to professional and local semantic networks. In the first scenario, the difference in network mean degrees would reflect the breadth of professional knowledge as opposed to the specificity and particularity of local knowledge; in the second one, the difference reflects the higher density of professional linkages—in both cases and in terms of network representation, this implies symmetric thresholding (i.e., if we use 2 for locals, we should use 2 for professionals), because we would otherwise lose the peculiarities of concept usage in professional knowledge.

By contrast, choosing a higher threshold for professional network than for local ones would filter away a great deal of structure to the point that some concepts in the local networks may erroneously appear to be more richly connected to their alters than in the professional network. This would additionally conflict with our initial assumption that locals draw on professional knowledge. While it may well be the case that some concepts are more elaborated in the local knowledge than in the professional one, we decided to leave this option for further research and focus instead on the assumption of inclusion of local networks into professional ones.

Illustration

We apply our approach to find and examine concepts local actor *a9* shares with professionals. We inspect concepts based on their similarity profile, because we want to understand what it means for two semantic networks to share concepts in terms of those concepts' alters. Comparing concepts, we take the following steps:
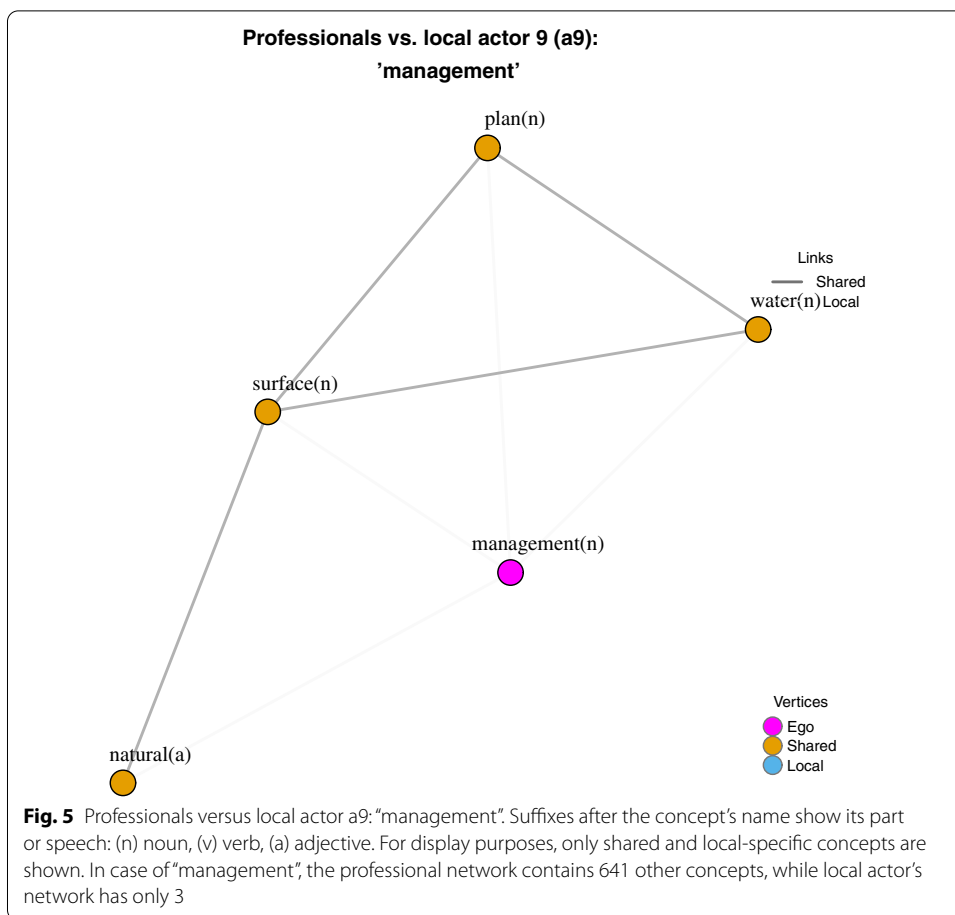
Fig. 4 Top 15% of the most central concepts for local actor a9 along with their vertex and link similarities to professionals

1. First, we create a list of common concepts: those concepts that both the professional and local actors use.
2. For each common concept *c*, we extract its immediate alters in professional and in local network. This yields concept-centered networks: $C_{\text{pro},c}$ and $C_{\text{loc},c}$
3. We calculate two metrics characterising similarity between the two concept-centered networks: vertex overlap and link overlap indices

Figure 4 shows vertex and link overlaps for the top 15% of the most central concepts in the local network *a9*. Let us examine two of them—*management(n)* and *plan(n)*. The concept "management" is interesting because it yields the highest similarity scores in both vertex and link overlaps. In other words, local actor *a9* expresses the same associations to this concept as professionals do, both in terms of the composition of alters and the links between them. The concept "plan", on the other hand, has relatively high link overlap but stands out with a relatively low vertex overlap score.

Concept "management"

Shared link "flood-management". Figure 5 displays the use of the concept *management* in the professional and a local networks. Let us start with a particular part of this cluster, the link "flood-management". The concept *management* often appears both in official documents and in interviews with locals and, indeed, is likely to play a pivotal role in both professional and local knowledge. There is a general understanding shared by both locals and professionals that floods cannot be totally eliminated and, therefore, should be properly managed to minimize their adverse impacts on people and the local economy.
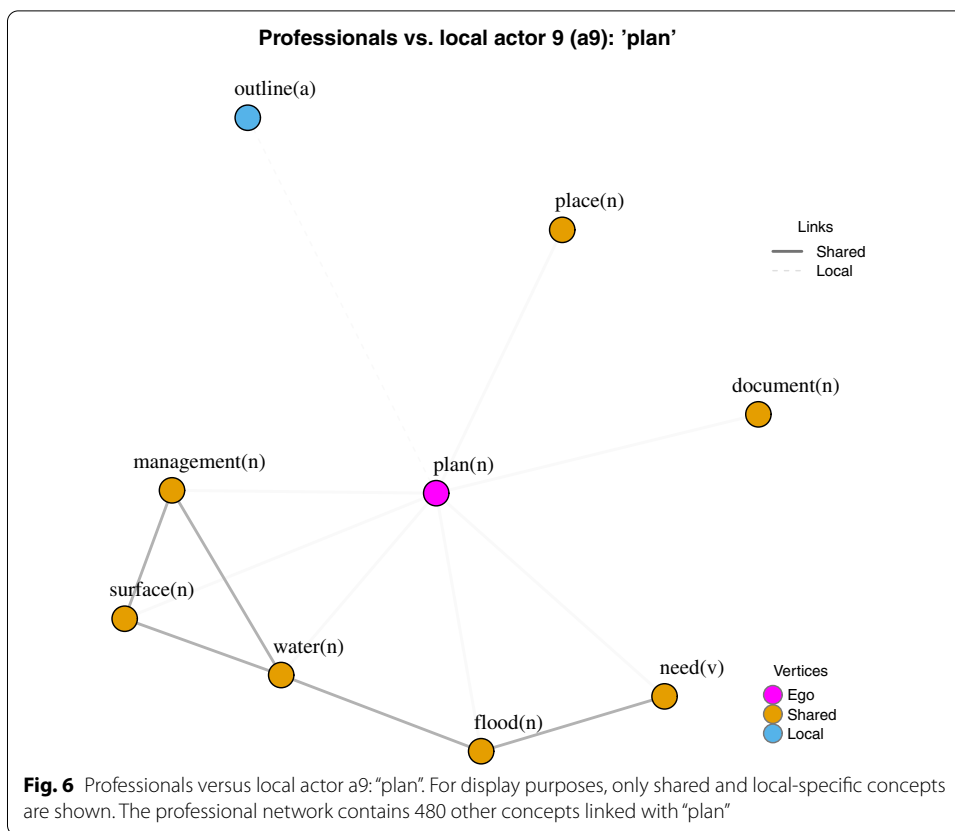
Fig. 5 Professionals versus local actor a9: "management". Suffixes after the concept's name show its part or speech: (n) noun, (v) verb, (a) adjective. For display purposes, only shared and local-specific concepts are shown. In case of "management", the professional network contains 641 other concepts, while local actor's network has only 3

> *I would say planning the most cost-effective solution for **flood management** surely that's the one isn't it? Nobody wants the flood but it will always come back to cost benefit of course so it's most cost-effective benefit for **flood management** that's close to the aim isn't it? (local actor **a9**)*

Shared clique "management-surface-plan-water". Professionals often use *surface water management plan* concepts to refer to an official document that coordinates and lead local management practice, with a special aim to minimize flood risk to properties:

> *In 2007 Telford & Wrekin Council were successful in a bid to create a **surface water management plan** under DEFRA's Integrated Urban Drainage pilot studies. The project was driven by the need to gain a better understanding of the surface water environment within its borough with a view to reducing the risk of flooding to existing and new properties through the development control process (professional text)*

The local actor, meanwhile, points out that ***surface water management plan*** is the main source of knowledge about the last flooding that many locals experienced themselves:

> *Most what I find with most documents related to flooding they're actually historical reports like the **surface water management plan**, the neighborhood plan, there may have been a report on the Priorslee balancing lake. (local actor **a9**)*

Fig. 6 Professionals versus local actor a9: "plan". For display purposes, only shared and local-specific concepts are shown. The professional network contains 480 other concepts linked with "plan"

Examining the use of "management" in both professional texts and local actor's interview shows that similarities in concept-centered networks correspond to similarities of this concept's meanings in verbal expressions. In other words, close reading helps confirming that comparing concept-centered networks at the level of vertex and link overlap indices corresponds to actual similarity of the usage.

Concept "plan"

Figure 6 shows the use of the concept *plan* in professional and local networks. *Plan* is indeed one of the core concepts in professional knowledge, in particular, *plan* is embedded into a clique with four other concepts some of which also appear in local actor's network: *surface-water-flood*. Meanwhile, we can also see that in the local network *plan* has several unique alters, most notably inside the triad *plan-neighborhood-connection* which does not appear in the professional network.

Shared link: plan-flood. "Plan" is another common concept in flood management vocabulary for professionals and locals alike. Professionals and locals use the concept "plan" when referring to documents that coordinate various stakeholders' flood management activities. Although both the locals and professionals share the idea of a document-driven approach to flood management, in practice they may refer to different levels of planning. For example, professionals most often refer to "flood risk management plans"—regional-level documents that orchestrate activities of stakeholders. In professional semantic network this is captured by the "surface-water-management-plan"

cluster of concepts. The regional level plan, however, does not directly relate to the locals. In fact, when speaking about plans, local actor refers to a "neighborhood plan"— a local document that regulates local development to protect the local drainage system from overload. This results in concepts "plan" and "neighborhood" being linked in the local network, yet never appearing next to each other in the professional network. Quotations below illustrate this difference in scale for professionals and local actor *a9*.

> *Flood risk management **plans** (FRMP) describe the risk of flooding from rivers, the sea, surface water, groundwater and reservoirs. FRMPs set out how risk management authorities will work together and with communities to manage flood and coastal risk over the next 6 years [...] Each EU member country must produce FRMPs as set out in the EU Floods Directive 2007. (professional texts)*

> *I suppose... the other one [issue] which isn't, perhaps, as major [a problem] but it [is] certainly significant for [the village], is the local developers. The planning permissions are granted on the understanding that certain flood mitigation steps will be taken... Developers are only allowed to develop in line with the **neighborhood plan**. (local actor **a9**)*

Additionally, the actor sometimes uses the concept *neighborhood plan* referring to some other local initiatives not even related to floods. The link to the concept *connection* indicates such use:
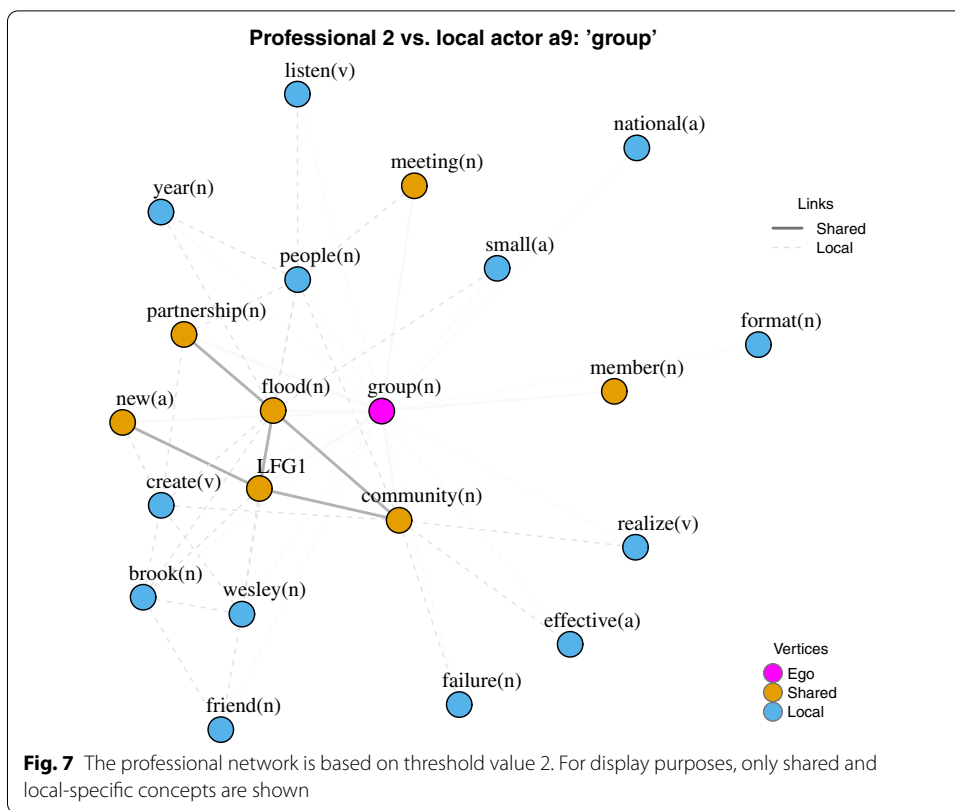
> *there was a very useful group which was called Shifnal Forward that actually brought all the community groups together with a purpose of a **neighborhood plan** of improving the railway **connection** to Birmingham, Wolverhampton and Shrewsbury for commuters business and various things like that (local actor **a9**)*

Dependence on the threshold

We mentioned earlier that threshold value affects network topology. In this section we explore this dependence. As a vehicle for the discussion, we use the concept *group*. In general, concept *group* has a very elaborated meaning in professional texts. There is, however, one aspect of meaning that locals and professionals share: *group* often refers to *action* or *community groups* of local people involved in risk management and representing local communities. This fact is clear when we look at professional network specification at the threshold of 2: the vertex *group* is connected both to the vertex *community* and to the vertex *action* (Fig. 7). Besides these two, the shared core of concepts around *group* also includes name of the local flood action group, *flood*, and *partnership*.

   Such intersection, in general, reflects one of the key ideas in current flood risk management in England: working in partnership with local activist groups. In fact, the very concepts *flood action group* or *community action group* have been coined and put forward by an NGO—the National Flood Forum—and local authorities. Locals, in general, appropriate such designation because their groups often have been established with the help of the National Flood Forum.

> *Continued support of **community groups** and forums as well as looking to broaden their understanding of surface water flooding (professional text)*

Fig. 7 The professional network is based on threshold value 2. For display purposes, only shared and local-specific concepts are shown
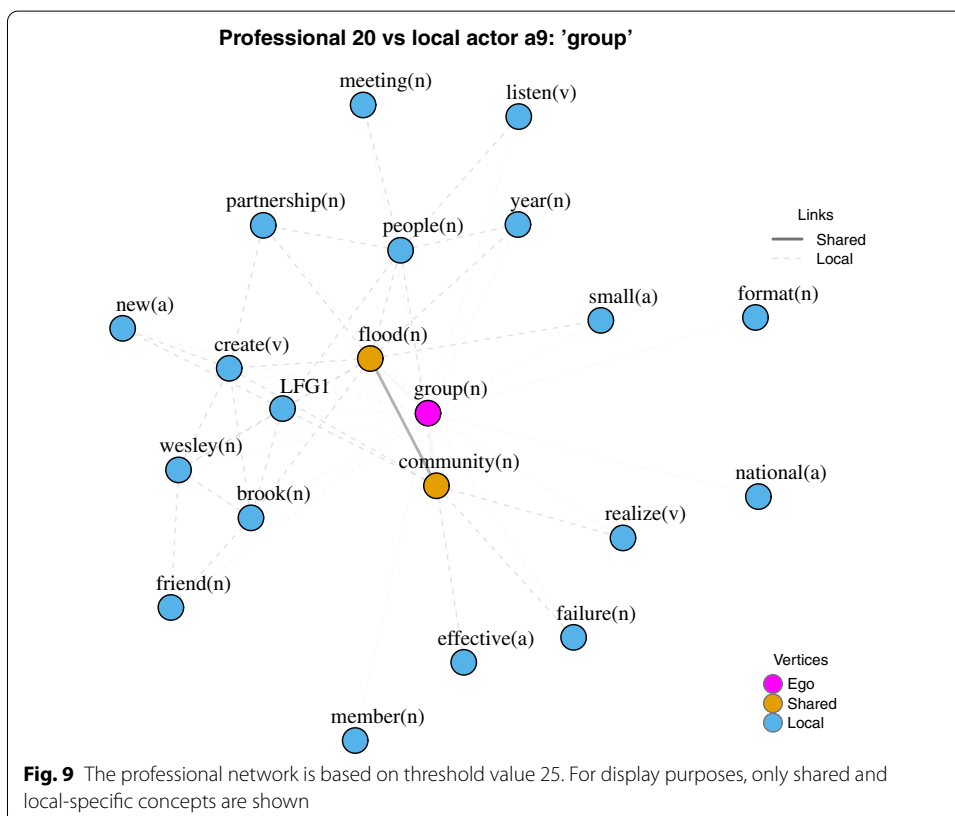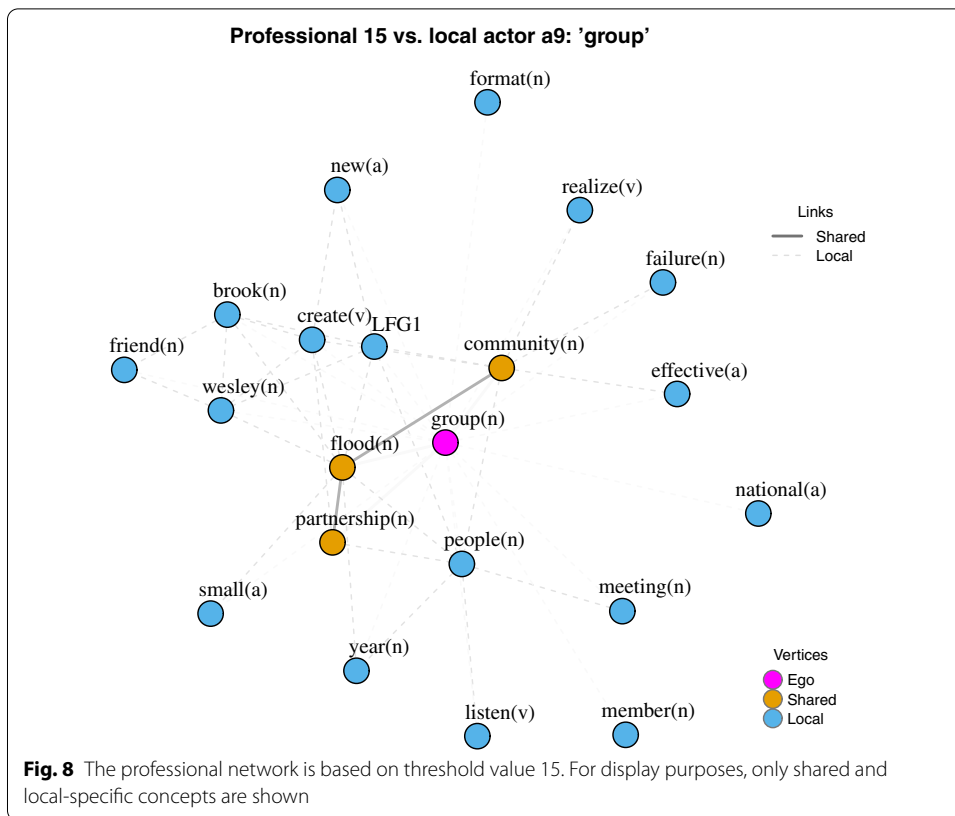
> *Try to have **action group** meetings at regular intervals (the multi-agency meetings will be less than the **action group** meetings) in that way the problems are kept to the fore even when flooding is the furthest thing on people's minds (professional text)*

More specifically, cluster *LFG1-flood-partnership-group* refers to a particular community flood group (the one that we studied). In fact, this is its official name: *LFG1 flood partnership group* proposed by the National Flood Forum.

At the threshold value of 15 (Fig. 8), the concept *group* retains only 3 concepts shared both by professionals and the local actor: *partnership*, *flood*, and *community*. The meaning of the concept *group* seems to be less specific for professionals, since the raised threshold filters the concept *LFG1* from their network. The reason may be that professionals use *flood* and *group* to denote any group in general, e.g. local groups of beneficiaries, working groups, etc. Manual check of professional texts featuring *group-partnership* co-occurrence without any reference to the research site *LFG1* supports this intuition:

> *Coastal **groups' partnerships** between maritime local authorities, the Environment Agency, and other key stakeholders develop SMPs and work together to implement them. (professional text)*

Finally, when the threshold is raised to 20 (Fig. 9), professional network retains only the cluster *group-flood-community*. The alters of this cluster also suggest very general use of the concept *group*, as if referring to broad categories of people or stakeholders.

Fig. 8 The professional network is based on threshold value 15. For display purposes, only shared and local-specific concepts are shown

Fig. 9 The professional network is based on threshold value 25. For display purposes, only shared and local-specific concepts are shown

> *There are three **groups** of people who form the primary beneficiaries of **flood** risk reduction through the provision of SuDS. (professional text)*
>
> *[...]to develop a strategic approach to drainage and **flood** management and receive reports from specific working **groups** and where applicable to inform other related working groups such as the Local Resilience Forum. (professional text)*

In general, different threshold values reveal different levels of abstraction used by professionals. When the threshold is low (2) we can see a lot of dyads and clusters referring to particular groups of people. As the threshold value increases, however, meanings become more general.

This simple exercise shows us that while what appears to be similar depends on the network construction protocol, the resulting variability still poses an interpretable material. Looking into how network pruning results in varying linkages between concepts sheds light on meaning structures and levels of abstraction employed by actors producing the text.

One straightforward implication here is that sensitivity analysis with pruning thresholds should become a necessary analytical routine for semantic networks. Besides that, however, we would argue that the value of working with a set of thresholds can go beyond a simple robustness check: we could also meaningfully interpret thresholds as levels of abstraction in text. As excerpts from professional documents above indicate, the same dyad (e.g. *flood-group*) can both refer to very specific and concrete groups, like *LFG1 local action groups* or to general categories like *groups of people* benefiting from flood management. Observing that the intersection between professional and local networks gradually narrows down from a dense and wide clique-like cluster to a single cluster provides insight into basic semantic dyads that, so to say, "pave the way" from professional to local knowledge.

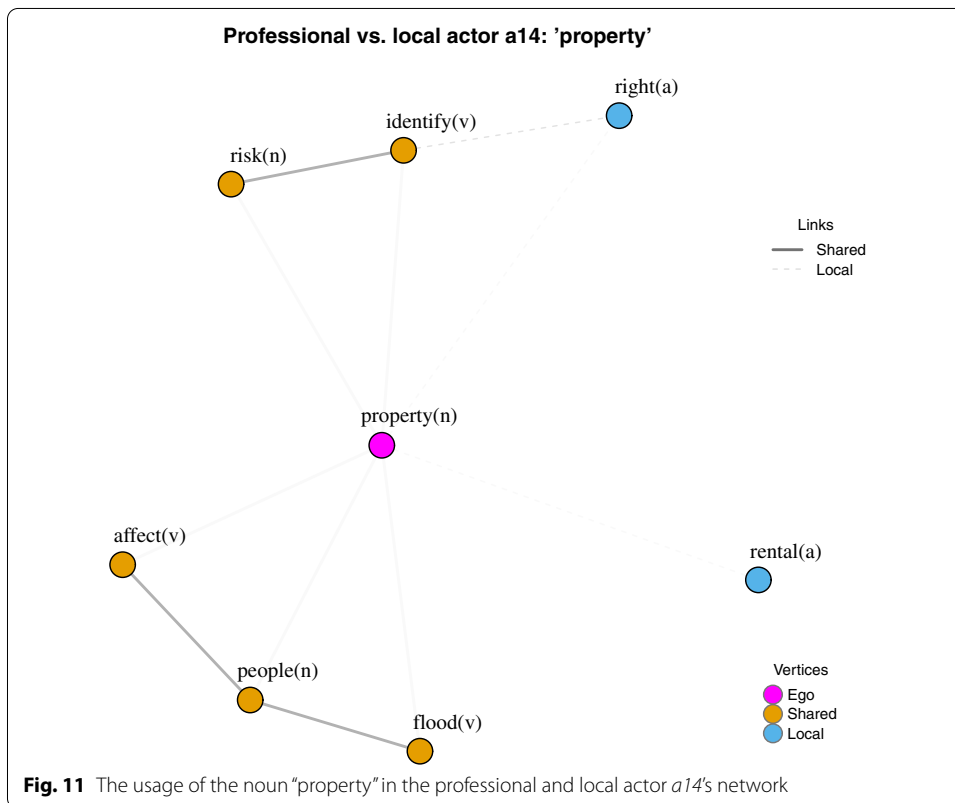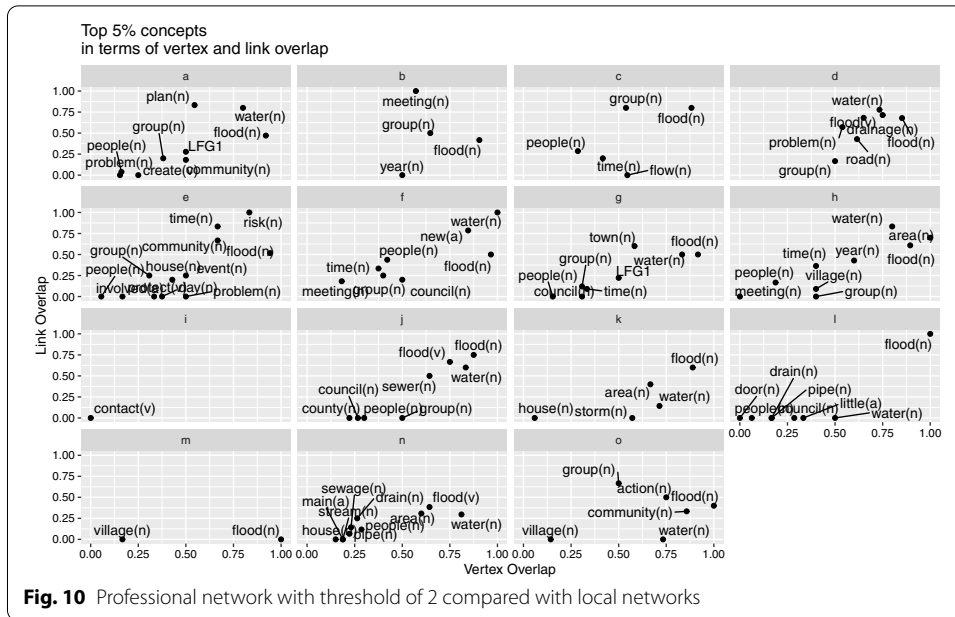### Comparing professional network with other local networks

Let us now repeat the procedure of looking for similar concept-centered networks for all the remaining local actors. Figure 10 compares local actors' networks against professional network based on the threshold of 2. Each panel presents top 10% of the most central concepts in a local actor's network, positioned by their vertex and link overlap scores. For illustrative purposes, we examine the concept *property* for actor *f* and concept *flood* for actor *n*.

Concept "property"

Figure 11 displays professional network and a local actor *a14's* network centered on the concept *property*. Shared part of alter-concepts include clusters *property-risk-identify* and *property-affect-flood-people*. The local actor has one unique concept: adjective *right*. Let us examine some parts of these clusters closer.

Shared: dyad "property-people". Shared dyad "property-people" hints at a tendency prevalent both in local and professional discourse that associates people with their property. For example, the local actor *a14* puts it:

> *So obviously there [were] inconvenience to the people that have flooded So you can't get in there to work. Like us we lost everything. So we literally couldn't go to*

Fig. 10 Professional network with threshold of 2 compared with local networks

Fig. 11 The usage of the noun "property" in the professional and local actor *a14*'s network

> *work . We lost our vehicles . We were homeless so the knock-on effect of that was that because 87 **people**, 87 **properties** were affected. (local actor **a14**)*

Professional texts also quite clearly state that protecting people means protecting their property.

> *Risk management authorities carefully prioritise maintenance activities to sections of rivers and the coast that provide the most benefit to **people** and **property**. (local actor **a14**)*

Shared: dyad "property-risk". Given that people are associated with their property, it is not surprising that both the local actor and professionals tend to bring in the concept *risk* when talking about property. For example, the local actor evaluates money received to mitigate flood damage in terms of properties at risk:

> *we were very fortunate that in January 2015 exactly 12 months after the money which the RFCC and the Environmental Agency wanted to give us according to the surface water management plan.. 87 **properties** were at **risk**.. in a 1-in-100 plus 30% climate change ADP so we got £520 000 pounds. (local actor **a14**)*

Meanwhile professionals describe floods in terms of risks they pose to properties.

> *This increase in flood water is believed to have increased flood levels at the junction of Meadow Road and High Street thereby increasing the **risk** and impacts of flooding to the **properties** at this location. (professional text)*

Shared: dyad "property-identify". Professionals and locals are both concerned with identifying properties at risk. Professional texts tend to use the verb identify speaking about zones in which properties reside.

> *Although the majority of Albrighton is **identified** as low risk of fluvial flooding. It must be **identified** that certain on Woodland Close and one **property** along Worthington Drive have been **identified** in Flood Zones 2 and 3. (professional text)*

The local actor, on the other hand, uses *identify* in a slightly different tone. First, he wants to identify properties vulnerable to floods because the number of properties at risk affects the funding their group receives from the county Council:

> *It might be that we have PLRs on seven the ones that are at higher risk one in a five but we did a Shropshire Council instigated a threshold survey to **identify** how many properties were at risk. So they sent out all these letters even to my neighbor's all up on the top of the hill. So it was done.. I'm telling you this about the threshold survey because that identify.. **So the number of properties that are that could be affected affects the funding**. (local actor **a14**)*

On top of that, he emphasizes that it is important that officials identify *right* properties, meaning properties that are identified as vulnerable by locals:

> *So when Shropshire council did this threshold survey we wanted them to do it to the **right properties** .. to the **properties** that we **identified** to them. (local actor **a14**)*

Fig. 12 The usage of the verb "to flood" in the professional and local actor *a7*'s network
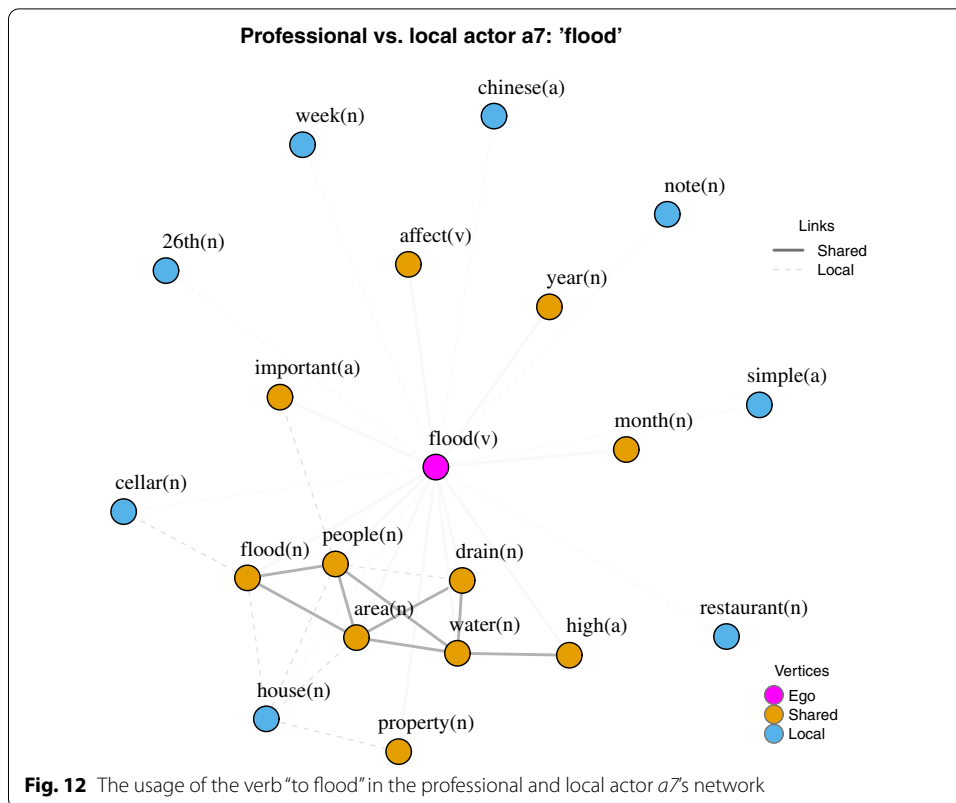
Concept "flood(v)"

Figure 12 displays the verb *to flood* for professionals and the local actor *n*. As always, we display only shared concepts and concepts unique to the local actor because of the large number of concepts linked with the verb *flood* in the professional network. We can see that while professionals and the local, indeed, share a lot of concepts linked with *flood*, the local actor has his unique links.

Shared: "flood-property". The shared core of concepts is not surprising and well fits general concerns about floods that professionals and locals share. Properties can get flooded and the use of the verb is almost identical for the local actor and professionals:
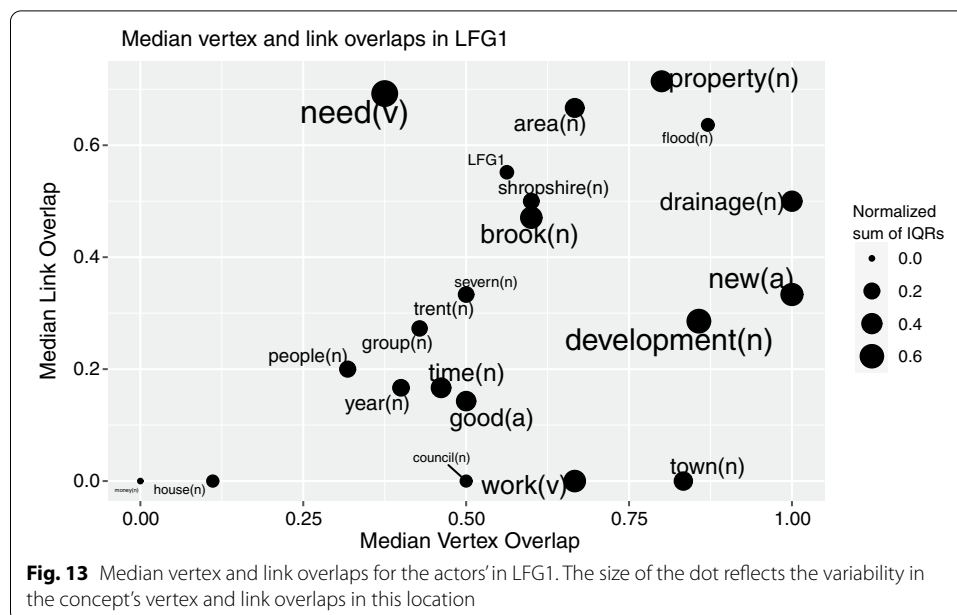
> no **properties** were identified as being **flooded** due to the LFG1 Brook overtopping its banks. (professional text)

> Woodland Close **flooded** on the 26s no **properties** affected. (local actor **a7**)

Note, however, that professional texts do not use the concept *house*, while for the local actor *property* and *house* are synonyms.

Non-shared: "flood-chinese". The local actor's network features many isolated concepts linked with the verb *flood*. One of them, the adjective *chinese* provides a curious detail on how local actors apprehend floods and its causes. Looking into interviews reveals that *chinese* refers to a *Chinese restaurant*, a potential location of a "natural stream" suspected to contribute to the recent flood in LFG2:

> flooded again about 4–5 months ago from that same .. it **flooded** the cottage next

Fig. 13 Median vertex and link overlaps for the actors' in LFG1. The size of the dot reflects the variability in the concept's vertex and link overlaps in this location
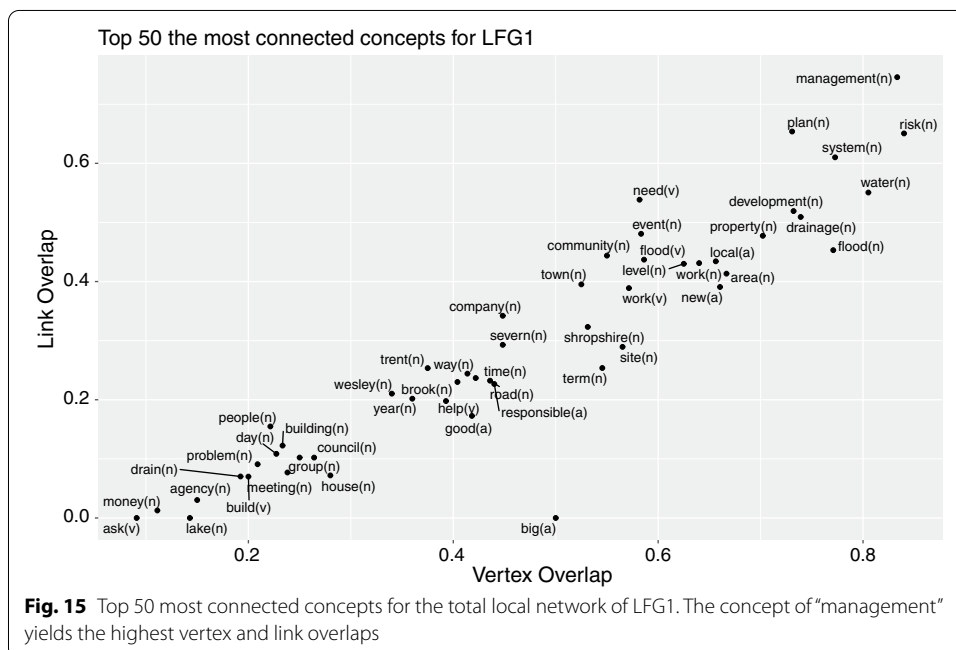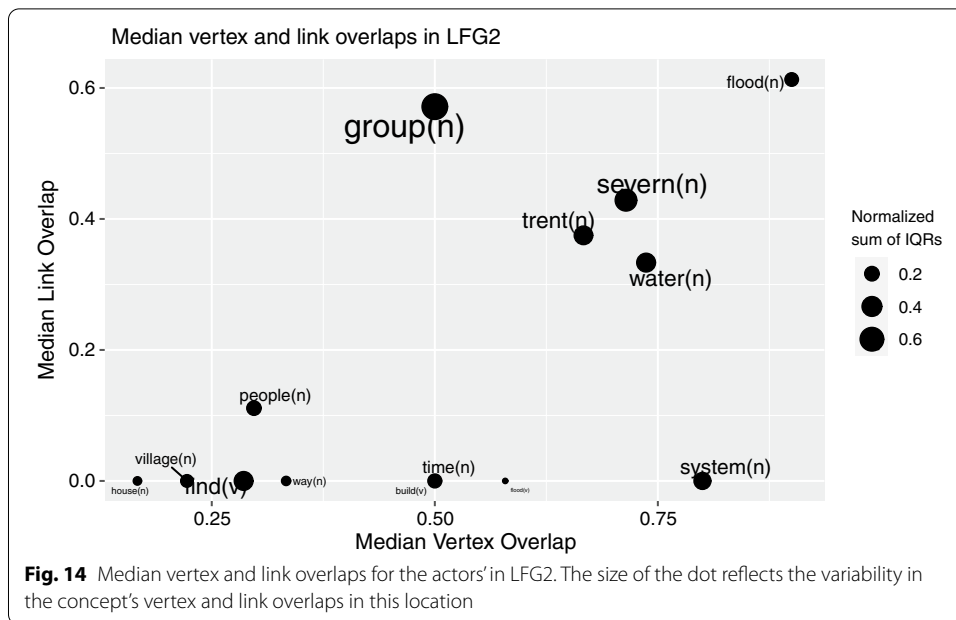
*to the **Chinese restaurant** .. it just came up and flooded so there's a question mark about that area it all comes back to the fact that we can't stop you connecting this housing development to it .. they have a right ... We know that there is another one somewhere [source of flooding] and we feel it hitting the concrete but where is the water going now. Now we do know an instance where if **flooded** what we know is a **Chinese** take away shop it flooded we now find that **there is a stream at the back of that shop** a natural stream. (local activist)*

### Comparing concepts within research locales

The last analytical step we perform is comparing concepts in their vertex and link overlaps between actors within research locales, i.e., LFG1 and LFG2. For example Fig. 13 shows median vertex and link overlaps for 18 concepts that all actors in LFG1 use. The size of the point reflects the associated variability—the sum of interquartile ranges for the concept's vertex and link overlaps. We may observe, for example, that both the concepts of *property* and *flood*, in general, have the same meaning for local actors as for professionals. The lower variability associated with the *flood*, however, suggests stronger consensus on the concept usage. Higher variability of the *property*, on the other hand, suggests that actors tend to differ in what aspects of property they discuss, thought these differences still more or less reside within the range of meanings provided in the professional texts.

The concept *money*, yielding zero vertex and link overlap indices, on the other hand, suggests that meaning of this concept for local actors is different from that for professionals.

Figure 14 shows the same distribution for LFG2. It is notable that the the number of shared concepts in LFG2 is lower than that in LFG1. We speculate that this might be because LFG1 was founded several years earlier than LFG2, so the difference in the pool of shared concepts perhaps reflects difference in the time spent to build these pools.

Fig. 14 Median vertex and link overlaps for the actors' in LFG2. The size of the dot reflects the variability in the concept's vertex and link overlaps in this location

Fig. 15 Top 50 most connected concepts for the total local network of LFG1. The concept of "management" yields the highest vertex and link overlaps

Comparing aggregates of local networks with the professional network can be an avenue for the future research, as it allows us to evaluate to what degree a whole union network of local knowledge is aligned with the professional network. For example, if we unite all the local networks from LFG1 site, we may observe that the concept of "management" yields the highest vertex and link similarity indices, paralleling that of the local actor 9 we discussed earlier (Fig. 15). However, if we examine the concept of "management" yielding the highest vertex and link overlap indices, we may observe that

Fig. 16 Top 50 most connected concepts for the total local network of LFG1. The concept of "management" yields the highest vertex and link overlaps
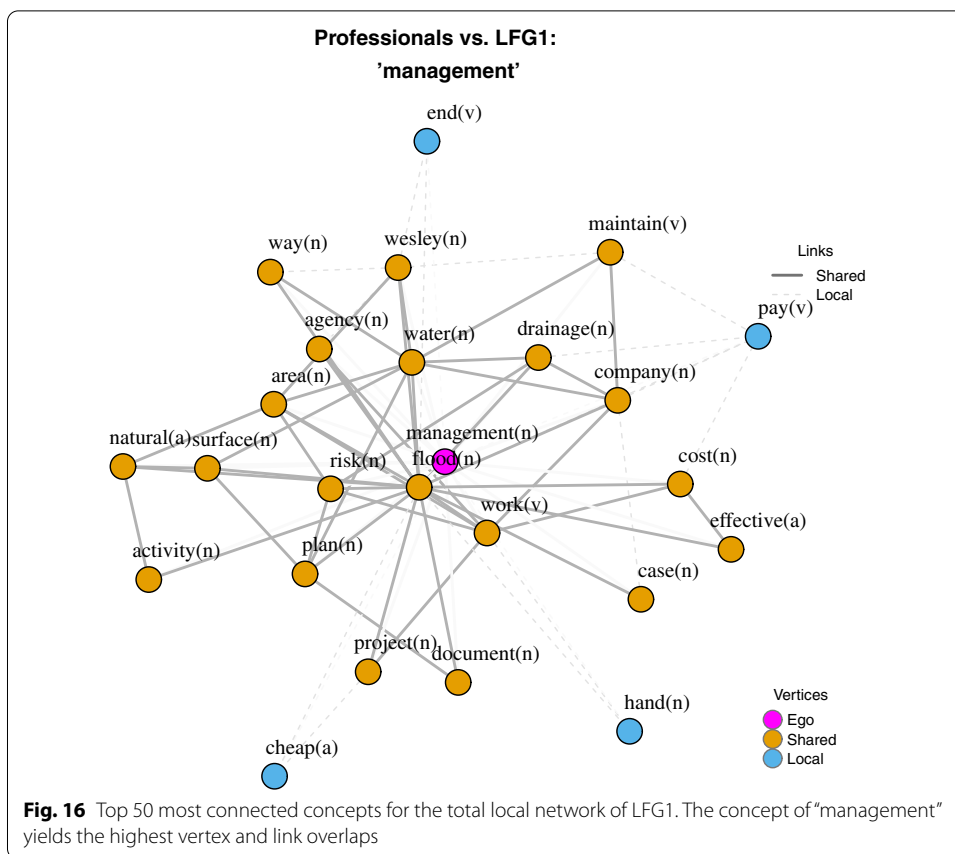
local actors in LFG1 collectively add to the idea of 'management" their own concepts as shown in Fig. 16.

Concluding remarks

This paper proposed a two-step concept-centered approach to compare semantic networks, where one network serves as a "golden standard" from which the other network selectively pulls semantic links. At the first step, we mapped all the shared concepts onto a two-dimensional space of overlap similarities of their alters and of links connecting these alters. The joint distribution of these indices highlighted concepts which potentially can give insight into selective appropriation of professional knowledge by local actors. At the second step, we visually inspect chosen concepts using a customized version of the Fruchterman-Reingold layout (Fruchterman and Reingold 1991), spatially separating shared and non-shared concepts.

We argued that while network comparison can happen at any level of analysis, in the case of semantic networks it is sensible to start with concept-centered networks. We suggest that researchers may gain deeper insight into how semantic networks emerge using the productive juxtaposition of quantitative and qualitative perspectives that this vantage brings together.

Finally, we want to discuss the shortcomings of the paper and the prospects of future research. Since we propose a method to compare networks, hypotheses testing is,

unfortunately, out of scope. We only show the utility of the method. Future research may apply the method to answer questions like "How do the meanings of the same terms in one knowledge system differ from those in another?" or "Do expert knowledge terms gain contest-specific meanings in local knowledge systems?"

From a technical side, we realize that the issue of threshold value needs to be addressed more principally, so it would be insightful to work with several symmetric thresholds for both professional and local networks. This implies that instead of working with one single network threshold researchers should embrace multiple network versions and explicitly incorporate this uncertainty into analysis. Finally, when analyzing shared and unique concepts across networks it deems necessary to look for qualitative explanations as to why local networks feature a lot of "singleton" vertices: concepts linked with the ego-concept but not with other alters.

Abbreviations
LFG1: Local flood group 1; LFG2: Local flood group 2.

Author's contributions
KP collected data. DM, CR, and NB developed the methodology. DM was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Availability of data and materials
Not applicable. The interviews have been collected under condition that the data will not be shared with third parties.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Sociology and Anthropology, Nazarbayev University, Nur-Sultan, Kazakhstan. [2]Centre Marc Bloch, CNRS, Humboldt Universität Berlin, Berlin, Germany. [3]Centre for German and European Studies, St. Petersburg State University, St. Petersburg, Russia.

References
Abbott JT, Austerweil JL, Griffiths TL (2015) Random walks on semantic networks can resemble optimal foraging. Psychol Rev 122(3):558–569. https://doi.org/10.1037/a0038693
Basov N, De Nooy W, Nenko A (2019) Local meaning structures: mixed-method sociosemantic network analysis. Am J Cult Sociol 9:376–417
Cantwell GT, Liu Y, Maier BF, Schwarze AC, Serván CA, Snyder J, St-Onge G (2020) Thresholding normally distributed data creates complex networks. Phys Rev E 101(6):062302
Carley K (1986) An approach for relating social structure to cognitive structure. J Math Sociol 12(2):137–189
Carley K, Newell A (1994) The nature of the social agent. J Math Sociol 19(4):221–262
Choobdar S, Ribeiro P, Bugla S, Silva F (2012) Comparison of co-authorship networks across scientific fields using motifs. In: IEEE/ACM international conference on advances in social networks analysis and mining. IEEE, pp 147–152
Dianati N (2016) Unwinding the hairball graph: pruning algorithms for weighted complex networks. Phys Rev E 93(1):012304
Fruchterman TM, Reingold EM (1991) Graph drawing by force-directed placement. Softw Pract Exp 21(11):1129–1164
Lee M, Marin JL (2015) Coding, counting and cultural cartography. Am J Cult Sociol 3:1–33. https://doi.org/10.1057/ajcs.2014.13
McEwen L, Jones O (2012) Building local/lay flood knowledges into community flood resilience planning after the July 2007 floods, Gloucestershire, UK. Hydrol Res 43(5):675–688
Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. Science 298(5594):824–827

Nye M, Tapsell S, Twigger-Ross C (2011) New social directions in UK flood risk management: moving towards flood risk citizenship? J Flood Risk Manag 4(4):288–297

Pržulj N (2007) Biological network comparison using graphlet degree distribution. Bioinformatics 23(2):177–183

Roth C, Cointet J-P (2010) Social and semantic coevolution in knowledge networks. Soc Netw 32(1):16–29

Wehn U, Rusca M, Evers J, Lanfranchi V (2015) Participation in flood risk management and the potential of citizen observatories: a governance analysis. Environ Sci Policy 48:225–236

Wijffels J (2019) Udpipe: tokenization, parts of speech tagging, lemmatization and dependency parsing with the 'UDPipe' 'NLP' toolkit. R package version 0.8.3. https://CRAN.R-project.org/package=udpipe

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.