Using Network Science to enhance the analysis of Delphi surveys' results in health settings

Carolina Paes de Faria[1]
carolina.p.de.faria@tecnico.ulisboa.pt
Instituto Superior Técnico, Lisbon, Portugal
November 2021

***Abstract*—Health Technology Assessment (HTA) systematically compares distinct aspects to evaluate health technologies and support decision-making processes concerning their use and financing. When choosing the evaluating aspects, it is common to use the Delphi technique to gather opinions from several stakeholders regarding their relevance. Although many authors implement Delphi surveys in this context, the study and abstraction of stakeholders' views lack exploration. In this work, we investigate innovative alternatives for the analysis of HTA Delphi surveys. To do so, we propose a framework based on Network Science tools. More specifically, the Louvain Community Detection algorithm is applied to cluster stakeholders according to their answers. We implement the framework to data from MEDI-VALUE and IMPACT-HTA projects, exploring two research questions: (1) verify the suitability of this approach for the analysis of Delphi surveys and (2) obtain novel and relevant information regarding stakeholders' opinions. The results suggest that the described framework is suitable for Delphi analysis, with the model originating communities with similar topologies when applied to different datasets. Additionally, the framework allows new insights to be obtained regarding stakeholders' opinions and how they relate. Communities found are typically not characterised by stakeholders' type and instead by the kind of answers, and people tend to consider criteria as relevant. The Triadic Closure is also verified for these networks meaning that if two stakeholders both agree with a third stakeholder, there is an increased likelihood that they will agree in the future.**

***Index Terms*—Health Technology Assessment, Delphi Technique, Proximity of Views, Analysis of Stakeholders, Network Science, Community Detection**

I. Introduction

The literature recognises that decision-making processes in the health sector are highly complex [1]. Not only are these processes influenced by several conflicting criteria, but stakeholders' perspectives and priorities frequently vary [1].

One particularly challenging process is resource allocation [2]. For the choice of using and financing health technologies, Health Technology Assessment (HTA) is used to support decision-makers by systematically evaluating several factors to compare health technologies [3, 4, 5, 2, 1].

Organisations are becoming more aware of the importance of defining and measuring the relevance of HTA aspects used to evaluate technologies as well as bringing the different stakeholders into the discussion [2, 3]. For that purpose, participatory approaches, such as the Delphi technique, are used to collect a vast number of opinions.

[1]This work was developed under the supervision of Prof. Mónica Duarte de Oliveira (monica.oliveira@tecnico.ulisboa.pt) and co-supervised of Prof. Francisco Correia dos Santos (franciscocsantos@tecnico.ulisboa.pt).

The Delphi technique consists in a series of surveys based on four principles - anonymity, iteration, controlled feedback, and statistical aggregation of responses [6]. This approach removes geographic and time challenges, allowing every stakeholder to participate while reducing the adverse sociological effects of group interactions [7, 6]. Although several authors conduct and analyse Delphi surveys' in the health context, the analysis and representation of opinions lack exploration.

Given the powerful applications of Network Science (NS) in other contexts, we propose an innovative framework based on NS and using Community Detection (CD) tools. This work has the goal of answering two main research questions (1) "Are NS tools suitable for analysing Delphi results?" and (2) "What information regarding stakeholders' views can be obtained?", each at one research stage.

This work focuses the analysis of Delphi results from the MEDI-VALUE HTA project. The answers express stakeholders' opinions on the relevance of HTA aspects. The surveys conducted consisted of two-rounds Web-Delphis.

In round 2 (R2), in MEDI-VALUE, 134 participants, divided into four groups (patients and citizens, healthcare professionals (HPro), buyers, policymakers and academics (BPA) and industry), evaluated 34 aspects using a 4-point relevance scale ("Critical", "Fundamental", "Complementary", "Irrelevant") plus a "Don't know/don't want to answer" option.

In MEDI-VALUE, aspects were evaluated for implantable medical devices (IMD) and biomarkers-based *in vitro* tests (BBIVT). We also analysed one Web-Delphi's results from the IMPACT-HTA project, for comparison. Thus, a total of three datasets was used. Only R2 results were studied.

This document is organized as follows. Section II introduces the main concepts and the relevant state-of-the-art. Section III presents the proposed framework and section IV describes details regarding its implementation. Results are illustrated and discussed in section V and the main conclusions are summarised in section VI.

II. Background and Related Work

A. Delphi surveys' results analysis

More important than the conduction of a survey is the data generated and its analysis and report [7]. Delphi originated data can be of many types, with the items' responses being commonly followed by comments. Its analysis can use both qualitative and quantitative methods [7, 8, 9], occurring between rounds, when providing participants with feedback and, in the end, more exhaustively. The items are this work's focus.

For the analysis of items' responses, many studies present statistical summaries for each item [9, 10, 11], such as frequency counts [12, 13] and range of scores [14]. Descriptive analyses are usually used to describe group agreement, mainly central tendency measures, i.e., mean, median and mode and level of dispersion, as standard deviation, coefficient of variation and interquartile range [7, 15, 9]. Additionally, the authors typically compare groups of stakeholders, using, for instance, Multivariate Analyses of Variance (MANOVA).

Besides the written narrative, results are commonly presented using tables of descriptive statistics [7] and the inclusion of graphics is recommended for easier visualisation [7].

Authors use different methodologies for results' analysis and report [16, 17, 15, 18, 10]. However, they do not differ much. Even though it is becoming more common to use various tools, including innovative charts and different statistical measures, there is room for deeper exploration.

B. *Possible alternatives*

Although current methods are effective, usually serving their purpose, they represent few of the possible approaches. For HTA Delphi surveys, the results report is of great importance to support decision-makers, usually in further meetings or conferences. We believe that providing them with new and different insights and adopting a new perspective is of great interest. With this work, we do not intend to minimise the currently practised approaches. Instead, we propose exploring and capitalising data unconventionally, using unique and powerful tools already proved immensely promising in other contexts. We want to understand if different and fresh information can be obtained, with the final goal of enhancing health decision-making processes.

Considering the state-of-the-art techniques in data analysis, there is one that stands out - Machine Learning (ML). People from all over the world are using it for the most various purposes [19], from academic and scientific research to customer experience improvement and business operations. ML techniques provide solid and effective data analysis, and their importance is currently undeniable.

Another emerging field is NS. In particular, the study of Complex Networks (CNs), large-scale graphs that can describe a variety of systems, capturing the spatial, topological, and functional relations of the data [19]. Much attention is now given to social networks, where nodes represent people or other entities and edges their relationships or interactions [20]. We believe these can be used as an inspiration for new abstractions representing peoples' interactions as agreement levels.

Complex Networks can present communities, meaning groups of nodes with many interconnecting edges. Nodes from different communities have typically relatively few edges interconnecting each other [19]. Figure 1 describes a schematic network with a strong community structure.

The use of ML robust analysis tools in structured connected data represented by CNs [19] results in Community Detection, the identification of the network's structure based on the interaction of its nodes [21]. Again, the possibilities are immense, and the applications suit different contexts. This strategy already applied in social networks is promising to group HTA stakeholders according to their opinions.

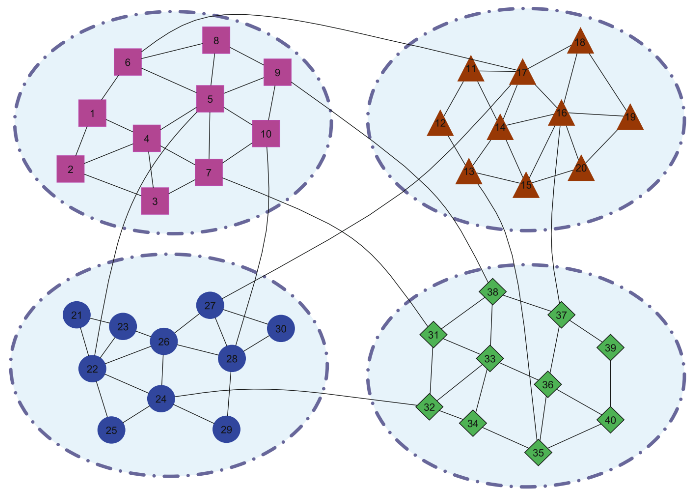


Fig. 1: Schematic of a clustered network, divided into four communities (source [19]).

We found no application of this strategy for HTA Delphi surveys analysis. However, we did find its use in other contexts with the goal of clustering stakeholders based on their views. For instance, [20], and [22] use this combination for social media networks' analysis. Other studies [23, 24, 25] focus on the identification of clusters of stakeholders, according to their argumentative similarities, however not in the health settings. These strategies are found to be suitable and powerful for similar purposes in different contexts. We want, in this work, to understand if they are also appropriate and as robust for clustering HTA stakeholders, according to their responses in Delphi surveys and better understand how they relate.

Thus, we propose to represent our data using CNs and cluster stakeholders according to their views, using a CD algorithm.

III. Proposed Framework

We apply NS, unconventionally analysing health Delphi results to find communities of similar stakeholders. For that, going back to the main research questions, we expect to (1) show that NS tools are suitable for analysing Delphi results and (2) explore the added information that can be obtained regarding stakeholders' views.

Based on the guidelines from [19], we propose a framework comprising five main steps, briefly illustrated in figure 2:

A. Pre-processing of the vector-based dataset;
B. Measurement of proximity and similarity of answers;
C. Conversion of the vector-based dataset into network-based data;
D. Application of the CD algorithm;
E. Visualisation and analysis of results.

A. *Pre-processing*

This step might involve different transformations such as scaling, normalisation, or cleaning [19]. It includes verifying

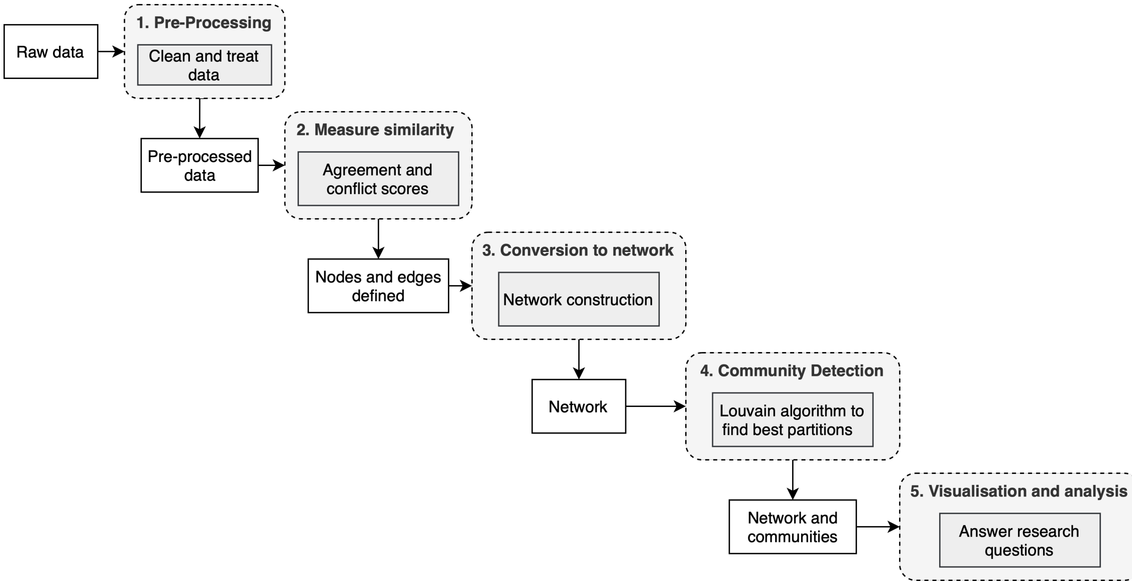


Fig. 2: Framework diagram.

if results are consistent with the study design, i.e., that all results were saved correctly and that no information was lost, including questions or participants related data. It is also necessary to transform and organise data to be further manipulated according to the methods meant to be applied.

B. Measurement of proximity and similarity of answers

The data analysed in this work is ordinal since it is originated from ordinal scales. These scales can suggest an ordering of people's opinions but not how distant they are [26]. There are several caveats regarding ordinal data analysis.

First, corresponding numbers to labels is often done but also a problem because it tries to quantify something categorical. Second, the use of parametric methods is frequent but controversial. Although some researchers such as [27, 28, 29] support their use, it is commonly defended that only non-parametric measures should be used [7], with the concern of otherwise getting incorrect results [30] or losing the meaning of conclusions [31]. Third, the interpretation of the midpoint item is also an issue, with research showing that people do not necessarily use it in the intended way [32].

With this in mind, we propose only to perform a comparison of answers with no use of parametric methods, setting the relationship between two stakeholders in one of four scenarios:

i. "Same answer", when both stakeholders choose the same scale point for a given aspect;
ii. "Same group", when the answers are not the same but the scale point chosen by both belong to the same group;
iii. "Opposite group", when the answers are not the same and belong to opposite groups;
iv. "Different group", when both answers neither belong to the same group nor the opposite group;

We also propose a definition for the groups. MEDI-VALUE's scale is not commonly described in the literature, and we found no guidelines on how to group items. Thus, the following option was assumed:

i. "Include with higher importance" - "Critical" and "Fundamental";
ii. "Include with lower importance" - "Complementary";
iii. "Do not include" - "Irrelevant";
iv. "No answer" - "Don't know / don't want to answer".

We consider "Include with higher importance" and "Do not include" as opposite groups. Interactions of "Same answer" or "Same group" are seen as agreement and "Opposite group" as conflict.

To define an agreement score between two stakeholders, we propose an adaptation of the approach described in [33], also used in [23] and [25]. For each pair of stakeholders, we consider two variables, an agreement variable A and a conflict variable C. For each aspect, if the relationship is considered an agreement, A adds up 1. On the other hand, if the relationship is considered a conflict, C adds 1. Finally, C is subtracted from A, as performed by [23] and [25].

C. Conversion into a network-based dataset

In order to apply a Community Detection algorithm, it is first necessary to transform the vector-based data into network-based data, i.e., transform each ID corresponding to a stakeholder into a node and define the relationships between the nodes, the edges.

Given a non-networked formatted set of N data points $X = \{x_1, ..., x_n\}$, it is possible to transform it into a network G, with the vertex set $V = \{v_1, ..., v_V\}$ and the edge set E, a subset of VxV using mapping procedure [19]:

$$g : X \rightarrow G = \langle V, E \rangle \tag{1}$$

Usually, as for this work, $X = V$, i.e., each data item from the original set corresponds to a vertex in the network. For N data items in X and V=$|V|$ the number of vertices, since no data reduction is conducted, V=N [19]. Finally, to obtain the set of edges E, a similarity function should be used followed by a network formation technique [19].

We use the described agreement score for the similarity function. For normalisation, aspects' scores are summed and the result divided by the number of aspects considered. For a pair of stakeholders, if their total score is higher than a given threshold, an edge is defined, representing their agreement.

D. Community Detection algorithm

Based on [34], we propose the use of the Louvain algorithm, first described in 2008 [35], to detect communities of similar stakeholders. This method merges nodes into communities in order to maximise modularity [34]. It stops when no merging results in a modularity increase.

Modularity is a measure commonly used to evaluate the quality of a partition [19, 35]. It is a scalar value between -1 and 1 measuring the density of intra-community edges when compared to inter-community edges [35, 34]. If e_{ij} is the fraction of edges connecting nodes from community i to nodes in community j, and $a_i = \sum_j e_{ij}$, the modularity Q is [36]:

$$Q = \sum_i (e_{ii} - a_i^2) \tag{2}$$

Modularity values close to 1 indicate a strong community structure [36]. It is rare to obtain high values in real networks, with typical ones ranging from 0.3 to 0.7. Note that when the whole network is a single community, the two terms are equal and therefore cancel each other, being the modularity equal to zero. In contrast, when each node constitutes a single community, the modularity is negative [36].

The Louvain algorithm is intuitive and of easy implementation. Also, it is better suited when the results are meant to be used as part of a broader analysis [34], which is aligned with this work's scope. It has an excellent computational time, with a linear complexity on typical and sparse data. Additionally, it is commonly mentioned in the literature as a powerful approach, outperforming other algorithms [37, 38, 39].

E. Visualisation and analysis of results

Both visualisation and analysis focus on the two research questions, each answered at one stage. In the first stage, networks and communities are analysed to understand how Delphi data is converted into network-based data and to find if there is a pattern of results for different inputs. This way, we can validate our framework. At this point, results from individual aspects and all aspects considered simultaneously are used. In the second stage, we analyse the networks and originated communities more detailedly, answering more specific questions for nine groups of related aspects and all aspects simultaneously.

Considering stage 1, we can compare the topology of the three networks. Since there is no previous report of using network analysis for Delphi, it is essential to investigate if there is a typical structure of these networks. Measures such as average degree and density are proposed.

In the context of social representations, the degree is the total number of a node's neighbours, i.e., the people with whom someone is connected [19]. In this work, we consider degree as the number of stakeholders an individual agrees with, with their agreement score overcoming the threshold.

Definition 1: **Density:** The density d of an undirected graph G, with n nodes and m edges, is given by [40] [41]:

$$d = \frac{\#Edges}{Max.\#Edges} = \frac{m}{\frac{n(n-1)}{2}} = \frac{2m}{n(n-1)} \tag{3}$$

The density of a graph is 0 when there are zero edges, and 1 for a complete graph [40]. In this context, a density of 1 means all stakeholders agree with each other. The lower the density, the less overall agreement there is.

In stage 2, we want to answer three main questions:

1) "What characterises the obtained communities?"
2) "Is there a clear division between communities?"
3) "Is the triadic closure property, from social networks, also verified in "agreement networks?"

Regarding question 1), we would expect stakeholders from the same type to behave similarly and communities to be defined by these types. However, if that is not the case, we can infer an added value from not using the stakeholder groups and rather understanding what clusters emerge. Furthermore, if communities do not match the stakeholder groups, we want to understand what defines them. To that end, we propose to:

i. Calculate the degree distribution per stakeholder type. If for a given type there is a variety in the degree distribution, it means similar stakeholders agree with different individuals;
ii. Explore the assortativity mixing, the tendency for nodes presenting similar attributes to be connected [42]. We propose the evaluation of assortativity regarding the stakeholder type to investigate whether similar stakeholders tend to be connected. The assortativity coefficient ranges between -1 and 1 [42]. A value closer to one corresponds to a strong tendency for nodes to connect to nodes with the same attribute. A value close to 0 means that there is no relevant association, and a value closer to -1 represents a network where there is heterophily [43];
iii. Measure the heterogeneity of communities, regarding stakeholders' type;
iv. Measure the distribution of answers per community. If communities are indeed not defined by stakeholders' type, we want to understand the properties characterising them.

For question 2), several inter-community edges mean that stakeholders agree with several individuals from other communities, indicating a higher susceptibility for opinion changes. We propose to analyse intra and inter-community edges and external ratio. According to [23], the external ratio is the number of inter-community edges as a proportion of total edges, allowing to measure the agreement between communities and how likely are individuals to change opinions.

Social Network theory shows that if A is connected to B and if B is connected to C, it is highly likely that A and C are also connected. This is called the triadic closure principle, which states that "*If two people in a social network have a friend in common, there is an increased likelihood that they will become friends themselves in the future.*" [44]. According to [36], a high level of transitivity or clustering coefficient is associated with the triadic closure. Both measures range from 0 to 1 [45, 46]. In question 3), using these measures, we want to investigate if this also happens in "agreement networks". Therefore, we can state that if A agrees with B and B agrees with C, A and C are more likely to agree. This way, we can learn more about HTA stakeholders' networks and compare them to typical social networks.

IV. Framework Implementation

This works' implementation was performed using Excel and Python. Excel was used for raw data (*.xlsx* files) pre-processing. All the code was implemented from scratch using Python. The code is publicly available in [47].

Python offers powerful packages which facilitate the construction and analysis of CN and implementation of CD algorithms. The *NetworkX* library allows the study of graphs and networks, and the *Community API* library implements the Louvain CD algorithm and allows further analysis.

For the establishment of the edges, a threshold of 0.6 was applied. This value was optimised by sensitivity analysis.

V. Key results

The results from each research stage are now presented.

A. Research stage 1 - Interpretation of results and suitability of the framework

We start by discussing the validation of the framework. To understand the transformation of raw data into network-type data, we first analysed isolated aspects from both projects and explored the topology of networks and communities.

The results are according to the expectations. Nodes and edges correctly represent stakeholders and their interactions, considering the answers given. The more diverse the stakeholders' opinions are, the fewer connections, and therefore edges, average degree, and density. The same happens for the number of communities, with more being detected when opinions are more diverse. We also analysed all aspects together, and conclusions were similar.

Although minor variations occur, the topology of the networks is similar. For the three datasets, where a similar number of stakeholders took part, the number of edges, the average degree, the density, and the community structure are generally similar. This finding is important since it shows that networks and communities' topologies are similar for identical setups but different participants, questions, and scales, supporting the use of NS and CD in this context.

B. Research stage 2 - Information extracted

After our frameworks' validation, at this stage, we explore which information concerning stakeholders' views and relationships can be extracted. Remember that the main questions are ”What characterises the obtained communities?”, ”Is there a clear division between communities?” and ”Is the triadic closure property also verified in this context?”.

Note that when applying CD tools, we are interested in strong communities. Hence, although the algorithm finds them, we are not concerned about clusters with few individuals. For further analysis, we will only consider the main communities with a considerable number of stakeholders.

For the sake of simplicity, because it would be too exhaustive to present results from the three datasets, the discussion will now focus on MEDI-VALUE's results. IMPACT-HTA analysis led to similar conclusions.

1) ” What characterises the obtained communities?”:

Initially, we want to verify what characterises communities. We would expect similar stakeholders to have similar opinions. Plus, Delphi results are usually analysed based on groups of stakeholders' types. Thus, we want to investigate if the communities match these groups. If not, our framework brings an added value to Delphi surveys' analysis since it starts with no pre-defined groups and instead investigates which clusters emerge. Furthermore, if the type of stakeholders does not characterise communities, we want to understand what does.

We started by analysing degree distribution. The results suggest that similar stakeholders do not necessarily share similar views. As an example, MEDI-VALUE IMD results from the group ”Value for the patient” are shown in figure 3.

In general, histograms have several bars representing a heterogeneity of degree within groups, suggesting that stakeholders from the same type are connected to different people, i.e., agree with people in different patterns. There are some variations, but they are small and justified. For example, probably because of the ”Value for the patient” scope, patients and citizens have fewer bars, meaning they have more similar views within their group. Overall, there seems to be a diversity of degree distribution per stakeholder type for all aspects.

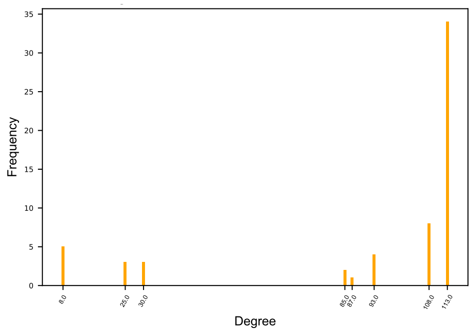
When we analysed all aspects together, we obtained similar results. With some minor differences, overall, stakeholders from the same group present different agreement patterns.

Considering the attribute assortativity coefficient, as shown in table I, for all groups, the values are close to zero. Hence, there is no higher likelihood for similar stakeholders, concerning their type, to be connected, meaning agree with each, than they would randomly. Again, values when evaluating all aspects simultaneously are similar, being close to zero.

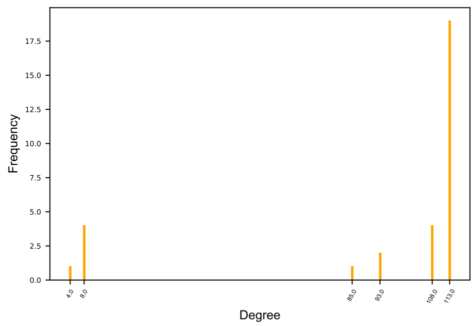
TABLE I: Attribute assortativity coefficient for MEDI-VALUE's groups of aspects.

Group	Implantable medical devices	Biomarkers-based *in vitro* tests
A	−0.009 081 6	−0.014 183 3
B	−0.008 387 0	−0.011 462 7
C	−0.011 360 8	−0.009 433 6
D	−0.011 562 3	−0.009 319 7
E	0.008 578 3	−0.022 345 9
F	−0.008 113 2	−0.010 572 8
G	−0.012 255 5	−0.009 010 4
H	−0.010 609 2	−0.012 615 2
I	−0.008 623 3	−0.011 457 0

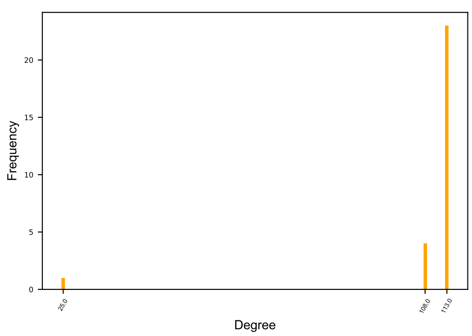
We also investigated the distribution of stakeholders per community. Overall, for groups and all aspects together, stakeholders are widely distributed across communities. When a given type is concentrated in a community, it is usually because all stakeholders are and not due to their type.



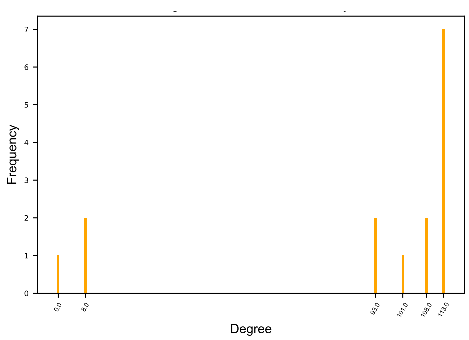
(a) HPro



(b) BPA



(c) Patients & Citizens



(d) Industry

Fig. 3: Degree distribution regarding IMD from MEDI-VALUE, for the Group A "Value for the patient".

Now we present the analysis for groups of related aspects in more detail.

Figure 4 presents a multipartite network for the MEDI-VALUE's Group A "Value for the patient" (IMD), showing the distribution of answers across the main communities and stakeholder groups. Figures 5 (a) and (b) present the distribution of stakeholders and types of answers per community.

These visualisation tools allow the reader to verify that no match exists between the communities and stakeholder groups. That can be observed by comparing the lines in the left and right side in figure 4 or looking at figure 5 (a). The majority of stakeholders belong to community 1, meaning that similar stakeholders gave similar answers, not because of their type but because the majority of all stakeholders gave similar answers ("Critical" or "Fundamental").

In figures 4 and 5 (b) it is also possible to observe that community 1 is mainly composed by individuals who answered "Critical" or "Fundamental", and a few "Complementary". Community 2 is more diverse but still characterised by "Critical" or "Fundamental", and community 3 only has "Don't know/don't want to answer" responses.

In general, considering MEDI-VALUE's Group A (IMD), communities are not characterised by stakeholders' type. Instead, the three main communities that emerge can be divided into types of answers. Communities 1 and 2 are characterised by "Critical" or "Fundamental" answers, with 2 being more diverse, and community 3 by "Don't know/don't want to answer ". Remember that aspects from Group A are related to the value of the technology to the patient. Thus, it is interesting to notice that, as expected, the majority of patients and citizens answered "Critical" or "Fundamental", a few "Complementary" answers were given, and none "Irrelevant" or "Don't know/don't want to answer".

These representations help us understand if stakeholders' types influence opinions, but they can also help decision-makers. This new visual representation of stakeholders' views allows to better visualise each stakeholder's opinions and how groups of opinions are defined. For example, for Group A, decision-makers can easily observe that most participants believe that the aspects should be included in IMD evaluation. Additionally, they can easily understand who is more likely not to answer (mainly healthcare professionals, industry and BPA) and who has a stronger opinion for the inclusion (patients and citizens). This way, by making these representations part of the report used to discuss Delphi results, the following steps may be easier to define and implement.

We now explore a less specific group related to the "Societal context of the adoption of the medical device". When the discussion concerns the value for the patient, opinions are somewhat straight, and people express similar views even if with some differences. However, for less simple aspects, as Group H's, the opinions are more diverse. Notice figures 6 and 7. Diversity can be observed by the smaller difference between the lines' width in figure 6 and by communities' composition, in figure 7 (b), which contrast with results from Group A.

Therefore, the presented representations can easily and quickly inform decision-makers on how dispersed and varied opinions are and how easy it can be to achieve a consensus on the inclusion of the considered aspects.

Results from other aspects' groups were similar, with some differences on the answers' variety, depending on the context of the aspects. More specific and non-conflict groups related, for instance, to patients' value and safety, led to less diverse opinions than costs or societal context groups.

Since participants were the same and answered both IMD and BBIVT surveys at close points in time, we compared these results. They followed a similar structure with stakeholders

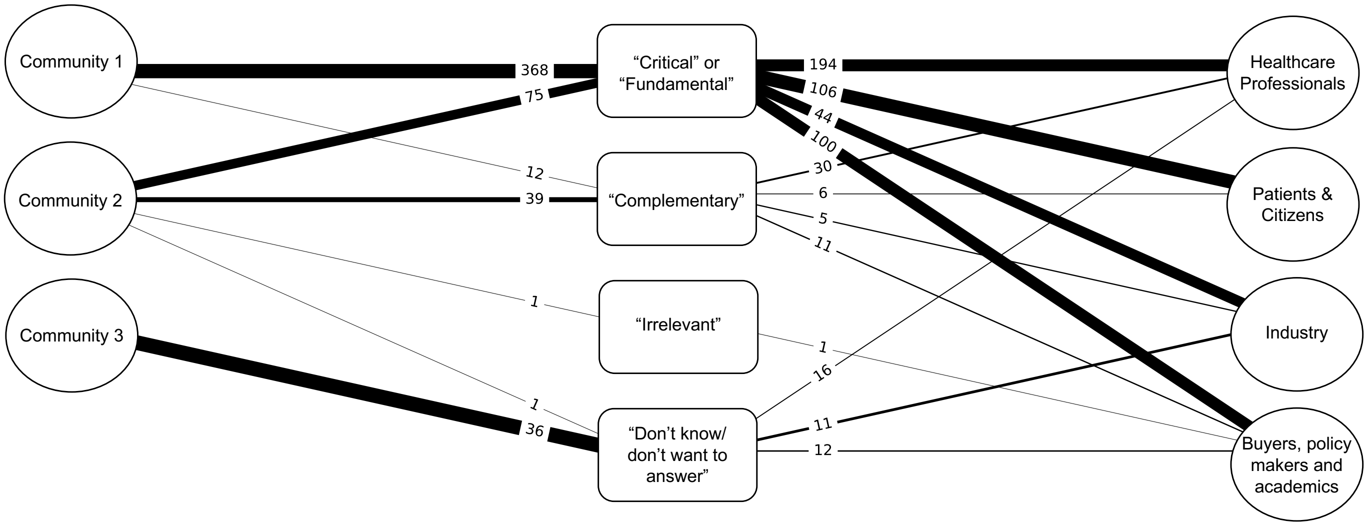
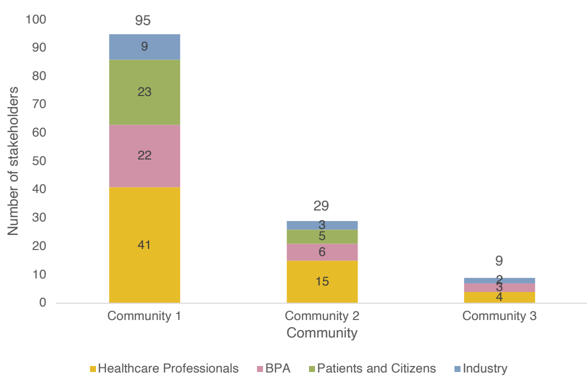
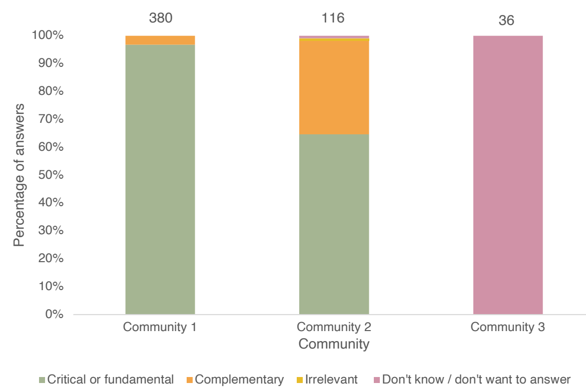


Fig. 4: Multipartite network, regarding MEDI-VALUE's Group A "Value for the patient" (IMD), showing the distribution of answers across communities and groups. Each line's value corresponds to the total number of answers, not stakeholders. This group comprises 4 aspects. The communities do not match stakeholders groups.



(a) Stakeholder types' distribution.



(b) Type of answers' distribution

Fig. 5: Distribution of stakeholders' type and answers, per community, for MEDI-VALUE's group "Value for the patient" (IMD). In (a) numbers regard stakeholders and in (b) answers. Stakeholders are widely distributed across the three communities. Community 1 is characterised by "Critical" or "Fundamental" answers. Community 2 is more diverse but still characterised by "Critical" or "Fundamental". Community 3 only has "Don't know/don't want to answer" responses.

scattered across communities and clusters being defined by the answers pattern. However, some differences, including the number and composition of communities, suggest that the technology being evaluated influences responses. As mentioned in [18], there are differences between drug therapies and medical devices impacting HTA. Therefore it is normal that differences exist. Also, although we count with a considerable number of participants, we are dealing with dozens of individuals not equally distributed between stakeholders groups. Thus, minor variations in opinions in one group corresponding to a few individuals might be perceived as greater.

These representations for all aspects led to similar conclusions but were less straightforward and insightful. We believe it is more recommended for decision-makers to discuss aspects group by group. For groups of related aspects, people are expected to have similar opinions regarding them. For example, if one stakeholder finds one aspect "Irrelevant", it is expected for him/her to share a similar view for the other aspects of the group. For that reason, opinions will be likely "polarised" for the answers' types for aspects' groups. However, for all aspects, people can have more diverse opinions. Moreover, stakeholders from the same community sharing similar views may find some aspects relevant but not others. This means that clusters are more likely to be characterised by patterns of answers. What matters is not only the type of answer but also which aspect the response regards.

Additionally, overall, a tendency for stakeholders to find aspects relevant was observed.

2) " Is there a clear division between communities?":

Now we evaluate if stakeholders have explicit, distinct opinions or agree with stakeholders from other communities.

We analysed the number of edges and found that results heavily depend on the threshold value and the number of evaluated aspects. The smaller the threshold value, the easier it is to consider an agreement and the higher the number of inter-community edges, meaning less defined communities. When few aspects are considered, communities are more defined because there are few possible scores. However, there is an

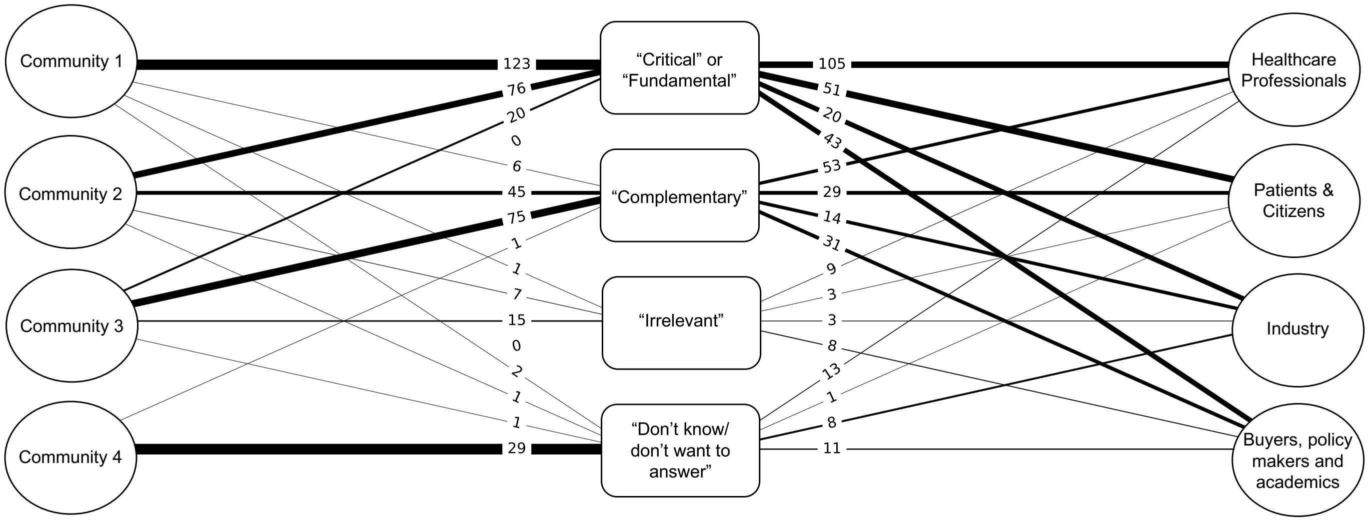
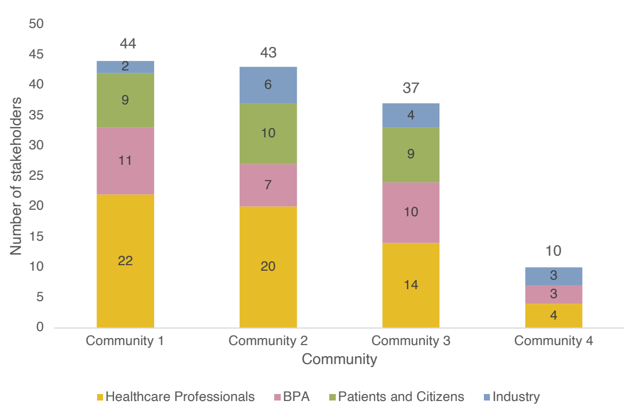
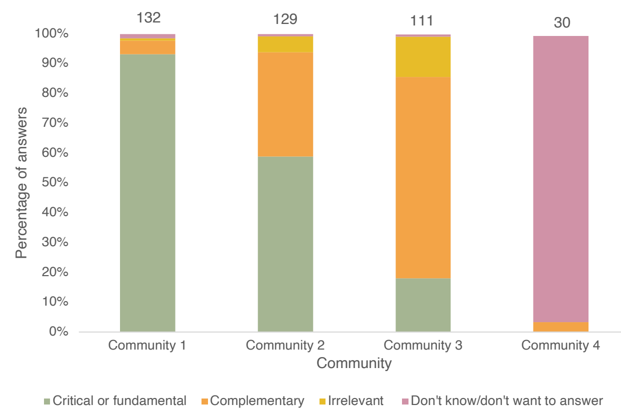


Fig. 6: Multipartite network, regarding MEDI-VALUE's Group H - "Societal context of the adoption of the medical device" (IMD), showing the distribution of answers across communities and stakeholder groups. The value in each line corresponds to the total number of answers, not stakeholders. Note that Group H comprises 3 aspects. The communities do not match stakeholders groups.



(a) Stakeholder types' distribution.



(b) Type of answers' distribution

Fig. 7: Distribution of stakeholders' type and answers, per community, regarding MEDI-VALUE's Group H - "Societal context of the adoption of the medical device" (IMD). In (a) numbers regard stakeholders and in (b) answers. Stakeholders are widely distributed across the four communities. Communities 1 and 2 are characterised by "Critical" or "Fundamental" answers, with the second presenting more diversity of answers. Community 3 is characterised by "Complementary" answers and community 4 by "Don't know/don't want to answer" answers.

increased probability that opinions start to be more diverse and more inter-community edges appear for more aspects.

Generally, communities seem not to be well defined. We obtained, for both groups of aspects and all aspects, several inter/intra ratios higher than 0. On the one hand, there are inter/intra ratios between 0 and 1, representing a considerable number of inter-community edges, i.e., stakeholders agreeing with people from other communities, suggesting a likelihood of reaching a consensus in the future. On the other hand, there are communities with ratio values much higher than 1, particularly for groups of aspects. These can represent good news, reflecting "weak" communities with individuals likely to change opinions and move to other communities. However, unfortunately, they can also indicate poor partition quality.

These findings suggest that communities are not clearly defined and that several stakeholders agree with individuals from other communities. Thus, people generally tend to have similar opinions at a given survey point, good news for achieving consensus. These conclusions should be made carefully for all aspects since numerous aspects are considered and therefore there are several possible response patterns.

3) "Is the triadic closure property also verified?":

Looking at table II, it is possible to notice that, for the majority of the groups of aspects, both transitivity and clustering coefficient values are close to 1. The results were similar when analysing all aspects together. Except for Group E, which has values around 0.6, probably due to the more diverse responses observed, in general terms, the triadic closure property of social networks is also verified in these "agreement networks".

This is an important finding. First, it defines a property of this new type of network, supporting the hypothesis that it is similar to social networks. Second, it describes a social behaviour highly relevant for the achievement of consensus.

One way of interpreting this property in social networks is "if two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future" [44]. If also verified in "agreement networks", we can infer that, *if two stakeholders both agree with a third stakeholder, there is an increased likelihood that they will agree in the future.* This property can be pivotal to describe the importance of common agreements and the influenced change of opinion.

TABLE II: Transitivity (Trans.), average clustering coefficient (Avg CC) and average degree (Avg Deg.) for MEDI-VALUE's groups of aspects.

Group	Implantable medical devices			Biomarkers-based *in vitro* tests		
	Trans.	Avg CC	Avg Deg.	Trans.	Avg CC	Avg Deg.
A	0.9477	0.9246	95.4627	0.9319	0.8449	63.0299
B	0.9729	0.9669	101.1493	0.9819	0.9180	81.2388
C	1.0000	0.9403	26.7600	1.0000	0.9478	28.3700
D	1.0000	0.9552	60.0400	1.0000	0.9776	61.1500
E	0.6860	0.6007	29.2388	0.7414	0.6813	30.6866
F	0.9949	0.9954	108.9851	0.9947	0.9746	96.0896
G	0.8130	0.7777	51.2090	0.8395	0.8089	52.7313
H	0.7827	0.7623	54.0900	0.8048	0.7785	52.0000
I	0.7484	0.7292	54.8955	0.8126	0.7895	55.3881

As previously mentioned, results for IMPACT-HTA were similar, according to its groups of aspects and scale used. Although the design and conditions of the surveys were distinct, the results were similar and analogous, corroborating our framework's suitability for analysing Delphi surveys' results.

VI. Conclusions

We explored the use of NS, particularly CD methods, to analyse Delphi surveys' results in health settings.

First, we proved that NS and CD are proper and promising to analyse HTA Delphi results, where nodes represent stakeholders and edges their agreement. We verified that the conversion of the original Delphi data into network-based data produced meaningful results. Additionally, we observed that networks and communities' topology were similar for both projects, meaning that the results from different surveys preserved a similar format. This way, we could demonstrate that Delphi "agreement networks" have a proper structure. We implemented our framework for two HTA projects, yet its employment is promising for other datasets and health contexts to explore stakeholders agreement networks.

Next, we explored new information that could be extracted. Our approach allowed us to verify that there is no significant match between stakeholder types and communities, which are mainly characterised by the type and patterns of answers given. When a group of stakeholders concentrated in one community, it was not because of their type but because all stakeholders were. One interesting detail is that there is a strong tendency for people to find all aspects highly relevant.

We also used different tools for analysing and representing the results and understand better the division and composition of communities, which are often not well defined. Results suggest a poor strength of communities which might facilitate consensus in the future. The results and used representations can serve decision-makers in conferences when discussing the inclusion of aspects, providing new and easy visualisation tools. Analysing groups of related aspects is recommended.

Furthermore, we concluded that the triadic closure property of social networks is also verified in "agreement networks". If two stakeholders both agree with a third stakeholder, there is an increased likelihood that they will also agree in the future.

One important finding is the high importance of the threshold value used to define agreement. Changing this value strongly changes results. Furthermore, we concluded that the triadic closure property of social networks is also verified in these "agreement networks". Thus, if two stakeholders both agree with a third stakeholder, there is an increased likelihood that they will also agree in the future.

Although the results were similar for the three datasets, there were differences in IMD and BBIVT results. This finding suggests that the context of the survey, meaning the type of technology being analysed, influences the results, which is not unlikely since these technologies are much different.

Our framework presents some limitations. In particular, the protocol, including the measurement and proximity of answers, was adapted from other studies. However, this framework has never been applied this way in this context. For this reason, it is challenging to compare and validate results. Moreover, several tasks were performed manually, which makes replicating the results more difficult and exhaustive than it would be with a more automated program.

There are several ways of improving our framework and exploring different approaches. These include performing more tests with different datasets, comparing results and also testing other approaches. Some of the proposals are overlapping communities algorithms, signed networks, and using Discourse Network Analysis to analyse the comments. Designing software to automatise the framework is also promising.

As a final remark, the possibilities are immense when discussing data analysis, including Delphi-originated data. There are currently several robust methodologies employed to study Delphi surveys' results that inform and support decision-makers. However, new and exciting approaches suitable in other contexts are constantly appearing. Thus, it was fascinating to find out at what level those can also be applied to HTA and Delphi surveys. With a great space for improvement, we believe we could complement the current Delphi analysis and provide decision-makers with fresh insights and visualisation tools, enhancing HTA processes. Finally, we hope this work inspires others to improve this framework and explore innovative approaches that might not be so obvious.

References

[1] J M Stratil et al. "Development of the WHO-INTEGRATE evidence-to-decision framework: an overview of systematic reviews of decision criteria for health decision-making". In: *Cost effectiveness and resource allocation* 18.1 (Feb. 2020).

[2] Ivett Jakab et al. "Potential Criteria for Frameworks to Support the Evaluation of Innovative Medicines in Upper Middle-Income Countries—A Systematic Literature Review on Value Frameworks and Multi-Criteria Decision Analyses". In: *Frontiers in Pharmacology* 11 (2020), p. 1203.

[3] Jackie Street et al. "Defining the role of the public in Health Technology Assessment (HTA) and HTA-informed decision-making processes". In: *International Journal of Technology Assessment in Health Care* 36.2 (2020), pp. 87–95.

[4] Stephanie Polus et al. "Health Technology Assessment of Public Health Interventions Published 2012 to 2016: An Analysis of Characteristics and Comparison of Methods". In: *International Journal of Technology Assessment in Health Care* 35.4 (2019), pp. 280–290.

[5] "Serviço Nacional de Saúde (SNS), INFARMED - Autoridade nacional do medicamento e produtos de saúde. *Avaliação de tecnologias de saúde*". In: (). https://www.infarmed.pt/web/infarmed/entidades/medicamentos-uso-humano/avaliacao-tecnologias-de-saude,accessed: 2021-10-23.

[6] Monica R Geist. "Using the Delphi method to engage stakeholders: A comparison of two studies". In: *Evaluation and Program Planning* 33.2 (2010), pp. 147–154.

[7] Ian Belton et al. "Improving the practical application of the Delphi method in group-based judgment: A six-step prescription for a well-founded and defensible process". In: *Technological Forecasting and Social Change* 147 (July 2019), pp. 72–82.

[8] Jenna M Evans et al. "A cognitive perspective on health systems integration: results of a Canadian Delphi study". In: *BMC Health Services Research* 14.1 (2014), p. 222.

[9] F Hasson, S Keeney, and H McKenna. "Research guidelines for the Delphi survey technique." eng. In: *Journal of advanced nursing* 32.4 (Oct. 2000), pp. 1008–1015.

[10] Anna Kearney et al. "Identifying research priorities for effective retention strategies in clinical trials". In: *TRIALS* 18 (Aug. 2017).

[11] Mary Falzarano and Genevieve Pinto Zipp. "Seeking consensus through the use of the Delphi technique in health sciences research." eng. In: *Journal of allied health* 42.2 (2013), pp. 99–105.

[12] Ivan R Diamond et al. "Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies". In: *Journal of Clinical Epidemiology* 67.4 (2014), pp. 401–409.

[13] Charlene R Weir et al. "Crossing the quality chasm: the role of information technology departments." eng. In: *American journal of medical quality : the official journal of the American College of Medical Quality* 21.6 (2006), pp. 382–393.

[14] Dyon Hoekstra et al. "Identifying and prioritising systematic review topics with public health stakeholders: A protocol for a modified Delphi study in Switzerland to inform future research agendas". In: *BMJ Open* 7.8 (Aug. 2017), e015500.

[15] Ângela Freitas et al. "Indicators for evaluating European population health: a Delphi selection process". In: *BMC Public Health* 18.1 (2018), p. 557.

[16] Edna Keeney et al. "Using a Modified Delphi Approach to Gain Consensus on Relevant Comparators in a Cost-Effectiveness Model: Application to Prostate Cancer Screening". In: *Pharmacoeconomics* (2021).

[17] Hanna Kallio, Pietila Anna-Maija, and Mari Kangasniemi. "Environmental responsibility in nursing in hospitals: A modified Delphi study of nurses' views". In: *Journal of clinical nursing* 29.21-22 (Nov. 2020), pp. 4045–4056. DOI: 10.1111/jocn.15429.

[18] Julie Polisena et al. "Health technology assessment methods guidelines for medical devices: how can we address the gaps? The international federation of medical and biological engineering perspective". In: *International journal of technology assessment in health care* 34.3 (2018), pp. 276–289.

[19] Thiago Christiano Silva and Liang Zhao. *Machine Learning in Complex Networks*. 1st ed. Springer International Publishing, 2016.

[20] K Sathiyakumari and M S Vijaya. "Community Detection Based on Girvan Newman Algorithm and Link Analysis of Social Media". In: *Digital connectivity - Social impact*. Vol. 679. Springer-Verlag Singapore PTE LTD, 2016, pp. 223–234.

[21] Zineb Felfli et al. "Community detection and unveiling of hierarchy in networks: a density-based clustering approach". In: *Applied Network Science* 4.1 (2019).

[22] Hongxu Chen et al. "People Opinion Topic Model: Opinion Based User Clustering in Social Networks". In: *Proceedings of the 26th International Conference on World Wide Web Companion*. WWW '17 Companion. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2017, pp. 1353–1359.

[23] Christina H. Buckton et al. "A discourse network analysis of UK newspaper coverage of the "sugar tax" debate before and after the announcement of the Soft Drinks Industry Levy". In: *BMC Public Health* 19.1 (2019), pp. 1–15.

[24] Shona Hilton et al. "Comparison of stakeholder coalitions across pricing policies designed to improve public health, as represented in UK newspapers: a discourse network analysis study". In: *The Lancet* 394 (2019), S52.

[25] Gillian Fergie et al. "Mapping discourse coalitions in the minimum unit pricing for alcohol debate: a discourse network analysis of UK newspaper coverage". In: *Addiction* 114.4 (2019), pp. 741–753.

[26] Ann Bowling. *Research Methods in Health Investigating Health and Health Services*. Open University Press, 2014.

[27] Geoff Norman. "Likert scales, levels of measurement and the "laws" of statistics". In: *Advances in Health Sciences Education* 15.5 (2010), pp. 625–632.

[28] Constantin Mircioiu and Jeffrey Atkinson. "A Comparison of Parametric and Non-Parametric Methods Applied to a Likert Scale". In: *Pharmacy* 5.2 (June 2017).

[29] Todd A DeWees et al. "Investigation Into the Effects of Using Normal Distribution Theory Methodology for Likert Scale Patient-Reported Outcome Data From Varying Underlying Distributions Including Floor/Ceiling Effects". In: *Value in health* 23.5 (May 2020), pp. 625–631.

[30] Susan Jamieson. "Likert scales: how to (ab)use them". In: *Medical Education* 38.12 (2004), pp. 1217–1218.

[31] Umesh Wadgave and Mahesh R Khairnar. "Parametric tests for Likert scale: For and against". In: *Asian Journal of Psychiatry* 24 (2016), pp. 67–68.

[32] Seung Youn Yonnie Chyung et al. "Evidence-Based Survey Design: The Use of a Midpoint on the Likert Scale". In: *Performance Improvement* 56.10 (2017), pp. 15–23.

[33] Philip Leifeld. "Discourse Network Analysis: Policy Debates as Dynamic Networks". In: *The Oxford Handbook of Political Networks* July 2016 (2017), pp. 301–325.

[34] Natalie R. Smith et al. "A Guide for Choosing Community Detection Algorithms in Social Network Studies: The Question Alignment Approach". In: *American Journal of Preventive Medicine* 59.4 (2020), pp. 597–605.

[35] Vincent D. Blondel et al. "Fast unfolding of communities in large networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10 (2008), pp. 1–12. arXiv: 0803.0476.

[36] Animesh Mukherjee et al. *Mukherjee, Choudhury, Peruani - Unknown - Dynamics On and Of Complex Networks , Volume 2*. Vol. 10. 2013, pp. 1–350.

[37] Zhao Yang, René Algesheimer, and Claudio J. Tessone. "A comparative analysis of community detection algorithms on artificial networks". In: *Scientific Reports* 6.August (2016). arXiv: 1608.00763.

[38] Karen Mkhitaryan, Josiane Mothe, and Mariam Haroutunian. "Detecting communities from networks : comparison of algorithms on real and synthetic networks". In: 26.3 (2019), pp. 231–267.

[39] Youngho Lee et al. "A comparison of network clustering algorithms in keyword network analysis: A case study with geography conference presentations". In: *International Journal of Geospatial and Environmental Research* 7.3 (2020), pp. 1–16.

[40] "NetworkX - Network analysis in Python, *networkx.classes.function.density*". In: (). https://networkx.org/documentation/stable/reference/generated/networkx.classes.function.density.html, accessed: 2021-08-31.

[41] *Baeldung, Determine Maximum Number of Edges in a Directed Graph*. https://www.baeldung.com/cs/graphs-max-number-of-edges, accessed: 2021-08-27.

[42] Matteo Cinelli et al. "Network constraints on the mixing patterns of binary node metadata". In: *Physical Review E* 102.6 (2020), pp. 1–22. arXiv: 1908.04588.

[43] Feng Shi. "Learn About Assortativity Coefficient in Python With Data From UK Faculty Dataset (2008)". In: *Learn About Assortativity Coefficient in Python With Data From UK Faculty Dataset (2008)* 2008 (2019).

[44] Morteza Aalabaf-Sabaghi. "Networks, Crowds and Markets: Reasoning about a Highly Connected World". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 175.4 (2012), pp. 1073–1073.

[45] *Transitivity*. http://users.dimi.uniud.it/~massimo.franceschet/teaching/datascience/network/transitivity.html, accessed: 2021-08-31. (Visited on 08/31/2021).

[46] *Data Science with Python, Triadic Closure (Clustering)*. https://necromuralist.github.io/data_science/posts/triadic-closure/, accessed: 2021-08-31. 2017.

[47] Carolina Paes de Faria. *GitHub, Thesis*. https://github.com/carolinapdfaria/Thesis. 2021.