TOOLS AND METHODS

Guidelines for carrying out a citation analysis: following evidence from production to use

William Hankey and Gabriel Pictet

Abstract

These guidelines describe the steps taken in Hankey and Pictet (2019) to carry out a citation analysis of a sample of the International Federation of Red Cross and Red Crescent Societies (IFRC) document base. The analysis followed evidence from production to use in order to assess what kind of evidence is produced and the degree to which it is taken up by other materials. These guidelines outline the steps taken to carry out the analysis and discusses theory in relevant parts. The first part covers document gathering and inclusion in the study, how to code the data, and briefly introduces the network analysis software used. The second part presents the approaches used, namely statistical analysis and network analysis. The former is used to analyse the dataset while the latter was used for metrics, visual analysis and analysis of the degree distribution. Finally, we discuss how the tools were used together to answer the research question. We hope these guidelines can help initiate a deeper discussion about how evidence is produced and used in the humanitarian aid and development sector.

Keywords: International Federation of Red Cross and Red Crescent Societies; citation analysis; evidence; humanitarian crises

Introduction

The importance of developing solid knowledge to respond to crises and emergencies, and the necessity of sharing lessons learnt has grounded humanitarian research agendas since the late 19th century (Davey et al. 2013: 1-2). The tools, methods and disciplines which form part of the humanitarian ecosystem have evolved with the emergence of new challenges and increased the efficacy of our actions. Yet evidence production remains a neglected area; few organisations know the extent of their knowledge base or whether evaluations and operational research findings are being capitalised upon. Furthermore, many practitioners are not trained to use evidence appropriately. As a result, evaluation and research efforts are regularly duplicated, poorly circulated and often misused (Clarke and Darcy 2014).

The evidence base in the sector is effectively a black box, a system whose inner workings are unknown, and which is defined solely by its inputs and outputs. In other words, we know what goes in and comes out (data and evidence), but not what happens in-between. Black boxing has a pragmatic aspect – one does not need to know the inner workings of a combustion engine to drive a car. But lack of knowledge about a black box means corrective action is difficult – without knowing how the engine works, one cannot fix a car. Gaining insight into how we produce and use our evidence is therefore vital if we are to put limited funding to good use, improve the uptake of research and, ultimately, be more effective in the field.

A common approach for studying the structure of scientific literature is citation analysis, a discipline related to network theory which examines how academic papers cite each other. In 2017, the International Federation of Red Cross and Red Crescent Societies (IFRC) carried out a citation analysis of a sample of its document base to understand how evidence is produced, circulated and used by the IFRC Secretariat and its National Societies. To carry out the analysis, we used an actor-network theory (ANT) perspective to trace referencing between documents and analysed the networks produced using statistical and network tools. The theory and results of the project are discussed by Hankey and Pictet (2019) in the previous volume of this journal. The study revolved around two questions: the first asks what strategies are employed to produce evidence, while the second asks how evidence is circulated and used to inform IFRC documents. These guidelines outline the steps taken to answer these questions. The first part discusses data gathering and wrangling, namely document gathering and inclusion in the study and how to score the data, and prepare for use in network analysis software, which we briefly introduce. The second part presents the approaches and tools employed, namely statistical analysis and network analysis. The former provides tools for analysing the dataset while the latter is a toolbox of metrics, tools for visual analysis and a method for analysing the degree distribution of a network. The final part examines how both approaches were used to answer the research question.

From documents to datasets – data gathering and wrangling

Document gathering and inclusion in the study

Documents were gathered in three ways, all of which rest on snowballing. First, we consulted colleagues and other contacts in the IFRC who could provide us with documents or share contacts. This included meeting National Society representatives during a workshop held in Geneva who provided us with contacts in their own organisations. If analysing only one organisation, consider finding a contact point in each relevant department instead. Second, we visited each participating organisation's website and downloaded any new documents we found. If any downloads were unavailable, we asked our contacts if they could locate them. Finally, we added documents to the sample as we encountered new documents in the reference lists of documents being analysed. These documents were acquired online or from our contacts.

A key step while gathering documents is to ensure that they meet the criteria for inclusion in the study. To develop the criteria, we started from a definition of evidence. A broad definition indicates that evidence results from the testing of hypotheses through theory and data (Snowden 2015). The Active Learning Network for Accountability and Performance (ALNAP) provides a more actionable definition: evidence is 'information that relates to a specific proposition, and which can be used to support or challenge that proposition' (Christoplos et al. 2017: 5) and that 'information only becomes evidence when it is related to a specific proposition' (Clarke and Darcy 2014: 7). In short, evidence is information that supports a specific proposition. In each case, evidence is the combination of a theory, an explanation of why something 'is', and the information that supports it. We follow the ALNAP definition above and consider 'evidence' as information that supports a specific proposition. Information and supporting propositions are often found in separate documents. Evaluation reports, for example, usually contain information on specific intervention outcomes which can be cited as evidence in a policy document that advances a given approach. Put simply, the evaluation report is an 'evidence document' and the policy is an 'evidence-based' document. The link between the two documents is the citation found in the policy document. An 'evidence-based document' is thus expected to cite documents that contain the evidence to support its claims. More specifically, a citation should: correctly reproduce and represent the content of a reference, make clear which statements references support, refer to the correct publication, and use a reliable source (Harzing 2002: 130-137). From this definition, evidence documents are documents we expect to produce evidence, namely research and evaluations. Evidence-based documents are documents which should use evidence, namely policies, frameworks, programme designs and advocacy documents. Since our analysis concerned the IFRC, we wanted documents produced or commissioned by the IFRC Secretariat, National Societies, or IFRC Reference Centres. We also wanted documents which focused on IFRC activities and topics. Finally, we wanted to concentrate on more recent research efforts and therefore limited ourselves to documents published within the last 5 years. Regarding exclusion criteria, we did not include financial documents, annual reports, HR reports, and other administrative documents. All criteria are summarised in Table 1.

Recording the data and creating the dataset

As we included documents in our database, we analysed them and recorded relevant data in our dataset. Aside from the features needed to construct the graphs, variables were drawn from the literature on social science and medical research evaluation, particularly mixed methods appraisal tools (see Long 2005 for a cogent example), and policy analysis. We will briefly present the variables and how the data was recorded.

Table 1. Document inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
Published within the last 5 years (2012 or later)	Draft document
Published or commissioned by the IFRC Secretariat, National Society or Reference Centre	Annual report
(Co)-authored by the IFRC Secretariat, National Society or IFRC Reference Centre	Financial report/audit/budget
Falls under a core IFRC activity or thematic sector	Presentation

The dataset is divided into three categories: document metadata, data on evidence production, and referencing data. Document metadata concerns data used to describe the documents in the dataset. We find data such as an assigned numeric ID, document title, author, publication year, its use of evidence, and so on. Assigning each document a numeric ID makes the identification of nodes much easier and facilitates the creation of edge lists, as will be seen below. Data on evidence production only concerns evidence documents. It asks questions about the transparency of evidence production in the document, such as whether it presents the methods used, any theory used, and if participatory methods were adopted. Finally, the referencing data groups data on the number of references and citations in the document, and which among these involve references to other documents in the dataset. This data, alongside the metadata, are used to produce the graphs. Table 2 provides details on each column in the dataset, including how to record the data. The assignment '0-9' indicates a numerical value, and 'a-z' indicates an alphanumerical value. The binary entry '0' or '1' indicates 'no' or 'yes', respectively. Finally, remaining columns have specified codes associated with them which are described in the table. Rows in bold, which cover evidence production, only apply to evidence documents - leave these cells blank when working with evidence-based documents. It is worth noting that, since each reference constitutes an observation, a new line is required to record each new reference being cited. As a result, some documents will be duplicated over several rows, with only the last four columns pointing to different references.
The final step before using the data is to run it through Google's OpenRefine (Ham, 2013). The latter is an open-access tool for cleaning data before processing it. It provides numerous functionalities which help pick out typos and other errors which easily creep into datasets. These functions are found in the drop-down menus next to column names.

Network analysis software

Once the dataset has been cleaned it can be imported to the network analysis software. Gephi (Bastian et al. 2009) and SocNetV (Kalamaras 2015) are open-access and therefore free to download. Both programmes were used since Gephi provides a more user-friendly platform for working with networks, but SocNetV provides additional analytic power. These programmes were also used since doing all these steps in R would involve a steep learning curve. The latter was nonetheless necessary to analyse network behaviour, as will be seen in the next part.

Table 2. Data codes and description. Rows in bold only apply to evidence documents

Data type	Variable name	Code	Description
Metadata	id	0-9	Numerical ID assigned to each document
	title	a-z	Title of the document
	year	0-9	Publication year of the document
	lead_author	a-z	Lead author of the document
	lead_org	a-z	Organisation to which the lead author is affiliated
	publisher	a-z	Publisher of the document
	doc_class	evidence	Is the document an evidence or an evidence-based piece?
		evidence-based	
	doc_type	research	What kind of document is this? An evidence document is research or an evaluation, and an evidence-based document is one of the other types.
		evaluation	
		policy	
		framework	
		advocacy	
		programme design	
	area_focus	DRR (Disaster Risk Reduction)	What is the main area covered by the document? You may adapt this typology to the areas covered by your organisation. 'all' indicates a general document, such as an overarching policy, which encompasses all your organisation's activities. The 'other' category can be used when no other category applies or you are unable to identify the area of focus of the document.
		health	
		social inclusion	
		livelihoods	
		CNVP (Culture of Non-Violence and Peace)	
		shelter	
		WASH (Water, Sanitation and Health)	
		migration	
		all	
		other	
Evidence production	**theory**	**0**	**Does the evidence document make use of a theoretical framework?**
		1	
	methods	**0**	**Does the evidence document include a section detailing the methods used and how the data was analysed?**
		1	
	approach	**quantitative**	**Does the study/evaluation adopt a qualitative, quantitative or mixed approach?**
		qualitative	
		mixed methods	
	participatory	**0**	**Does the evidence document take a participatory approach in any part of the research process?**
		1	
Referencing	citations	0-9	Total number of citations in the document
	references	0-9	Number of references in the reference list
	org_references	0-9	Number of IFRC references in the reference list
	org_title	a-z	Title of the IFRC document being referenced
	org_id	0-9	Index number of the referenced document
	org_citations	0-9	Number of citations to the reference

Mathematically speaking, a graph is composed of a set of nodes and a set of edges. We therefore need to develop a list of nodes and another of edges. The node list contains all the nodes and any attributes we want to assign to them (author, publication year, document type, etc.). The edge list contains all the connections between nodes (references) and their weight (number of citations), which determines the thickness of the edge. An important distinction is made here: edges represent references, weights represent citations. The easiest way to build each list is to copy columns in a new workbook and save the resulting node or edge list as a Comma Separated Value (CSV) file. At a minimum, the node list should include of the first and second columns ('id' and 'title' columns) and the class of the document ('doc_class'). Any other column can then be added as attribute. Finally, the 'title' column should be renamed 'label'. The edge list needs the first column ('id'), and the columns containing the reference ID and number of times it is citing the reference ('org_id' and 'org_citations'). The 'id' column should be renamed 'Source', 'org_id' should be renamed 'Target', and 'org_citations' should be renamed 'Weight'. Finally, remove any rows where there are no citations, leaving only rows where a link is made between two nodes.

The first step is to import each CSV file into Gephi, starting with the node list (File > Import spreadsheet). Keep the default options for importing and, on the final window after importing the data (named 'Import report'), select the option to append the data to the existing workspace. Repeat the same process with the edge list, again appending the data to the existing workspace. Once the data is imported, Gephi provides a simple interface for changing the appearance (top left) and layout (bottom left) of the graph. For the layout, we chose the Yifan Hu algorithm with the optimal distance set to 90, relative strength to 0.4, initial step size to 19, and step ratio to 0.9. This algorithm places higher degree nodes closer to the centre with isolated nodes lying around the periphery and allows for a rapid identification of important clusters and nodes. The Context window (top right) provides a summary of the graph and will allow you to confirm that the correct number of nodes and edges have been imported. To the right of the interface, the Statistics tab provides a quick way of calculating several network-level metrics, while the Filters tab provides a way of masking vertices and edges based on their attributes or metrics. Once network metrics have been calculated using the Statistics tab, vertex-level metrics can be found in the Data Laboratory tab, along the top ribbon. The latter presents node and edge data in spreadsheet format alongside vertex-level metrics which have been calculated through the Statistics tab. Finally, the graph can be saved as a PNG or PDF file by selecting the Preview tab at the top. Subgraphs, where nodes or edges have been filtered, can also be saved. For a more complete introduction to Gephi, see Ognyanova (2012).

To work with our graph in SocNetV we exported it as a Pajek file. This is done through the File tab (File > Export > Graph file, use the .net file extension). Using the Pajek file in SocNetV is simply a case of opening the file. Check there are no errors by examining the number of nodes and edges, which are displayed in the top right. SocNetV has similar capabilities to Gephi but provides more metrics and functions. All metrics can be accessed from the 'Analyze' tab at the top of the main window. If you wish to keep your analysis simple, you can keep skip using SocNetV and use only Gephi which.

Tools of the trade – how to analyse the data

Having described the software and how to import the data, we will discuss the three approaches used to analyse the data, namely, statistical analysis of the dataset, network metrics, and visual analysis of the graphs. Finally, we also describe how to analyse the presence of a power-law in the degree distribution. Aside from the statistical analysis, all tools are drawn from the literature on network theory.

Statistical analysis of the dataset

The dataset is the easiest element to analyse in the study since it can be done with minimal knowledge of Excel functions. We used the dataset to calculate proportions, means and other simple statistics. This included calculating the number of documents per author, the proportion of evidence documents that use theory, and the proportion of documents which use IFRC references, for instance. A statistical analysis was also carried out to calculate the probability of edge occurrence between document classes, types and areas.

Using and interpreting network metrics

Many network metrics are based on the concept of path, which is a sequence of nodes and edges between any two nodes in the network. The distance between two nodes is generally calculated as the number of edges between them. When implemented, metrics which use path length will always calculate the shortest path between nodes. These metrics therefore depict best-case scenarios.

Metrics are available at the vertex level or the structural level (the overall network). In our study, the former provided insight into a node's standing in the network, while the latter provided a summary of each network and some indications of their connectivity and performance. We used structural metrics to compare the different organisations in the study. The vertex-level and structural metrics used in the study and their actor-network interpretation are listed in Tables 3a and 3b, respectively.

Table 3a. Vertex-level metric definitions and actor-network interpretation

NA Metric	Definition (Newman, 2010)	Interpretation in ANT citation networks
Degree	The degree of a vertex is the number of edges connected to it. It gives a measure of how connected a vertex is to others in the network.	It measures how much a document cites (in-degree) or the number of times it is cited (out-degree) and provides a crude measure of whether a document is well-informed or influential, respectively.

Betweenness	Betweenness measures the extent to which a vertex lies on the paths between other vertices. It is a guide to the influence vertices have over the flow of information between others.	Documents with high betweenness are important in bridging groups of documents and exchanging new information across them. Removal of these documents will disrupt the structure of the network most as they lie on the largest number of paths between groups. As betweenness rests on a vertex having an out-degree, only documents which are cited will score on this metric.
Closeness centrality	Closeness centrality measures the mean distance from a vertex to other vertices. High closeness centrality indicates better access to information at other vertices or more direct influence on other vertices. Since it takes into account all vertices, we used a variant of the metric called information range closeness centrality which discards vertices with no degree.	As closeness centrality is based on in-degree, it provides a rough estimate of how much a document will draw in information, knowledge and evidence from surrounding texts.
Clustering coefficient	The clustering coefficient is the average probability that two neighbours of a vertex are themselves neighbours and measures how complete a vertex's neighbourhood is.	It measures the extent to which documents will use the same references.

Table 3b. Structural metrics definitions and actor-network interpretation

NA Metric	Definition (Newman, 2010)	Interpretation in ANT citation networks
Average degree	Average degree calculates the mean degree of vertices in a network and represents the how well connected the average vertex is.	It represents the average number of times documents will cite (in-degree) or be cited (out-degree) by other documents in the network.
Average path length	The average path length measures the mean number of edges along the shortest paths between any two vertices in the network. It measures the efficiency of flows in a network.	It measures how far, on average, any piece of information or evidence from one document can travel to any other connected document in the network.
Density	The density of a graph is the fraction of maximum possible edges in a graph. Maximum density is 1 (all possible ties are present), the minimal density is 0.	It measures the extent to which documents are citing each other relative to the maximum number of citations possible. A maximum value of 1 would be undesirable as only relevant citations need to be made between texts.
Diameter	The diameter of a graph is the length of the longest calculated shortest path between any pair of vertices in the network for which a path actually exists.	It provides a rough measure on how far information or evidence can travel across the network.

Average clustering coefficient	The average clustering coefficient calculates the mean clustering coefficient of all vertices in a network. It measures the extent to which vertices will form highly connected groups.	It denotes how much one can expect documents to share references across the network.
Modularity	Modularity measures the tendency of vertices with similar properties to connect. It is strictly less than 1, takes positive values if there are more edges between vertices of the same type than we would expect by chance, and negative ones if there are less. In other words, it is a measure of how structured connections in the network are.	It measures the extent to which texts will cite across document types and areas of specialty. A higher value indicates more referencing occurs across categories and therefore that there is more cross-fertilisation between domains.

While a single metric will provide some insight into a node's performance, interpreting them together provides better understanding of its role in the network. Table 4 summarises how centrality metrics are interpreted together.

Table 4. Interpreting centrality metrics

	Low degree	Low closeness	Low betweenness
High degree		The node is embedded in a cluster that is far from the rest of the network.	The node's connections are redundant, flows bypass it.
High closeness	The node is tied to other important/ active nodes.		The node is in a close-knit group with other nodes. There may be multiple key paths in the network.
High betweenness	The node's few ties are crucial to the network.	The node monopolises ties from a small number of nodes to many others.	

Visual analysis of a graph with Gephi

Gephi provides numerous functions for facilitating visual analyses. We started by colouring and resizing nodes according to different attributes and metrics. The aim of this step is to identify any patterns in how nodes cluster together. The layout algorithm should place connected nodes closer together, do clusters of nodes appear to share common attributes? Once an intuition for the network was developed, we used Gephi's built-in filtering functions to analyse subgraphs in greater detail. A subgraph is a graph embedded within in a larger graph. For example, policies and their references form a subgraph, as do documents produced in a given year. This allowed us to gather data (number of nodes and edges, for example) on different subgroups in the network and examine how they interact. The 'doc_class' attribute, for instance, was used with the MASK filter to count edges between and among evidence and evidence-based documents. From this data, we developed tables containing the number of edges between document types, areas, and other attributes, to analyse any relations in greater detail. The filters are accessed through the Library folder in the Filters tab (Library >

Attributes > Intra Edges / Inter Edges / Partition, and Library > Operators > MASK). The filters we used are summarised in Table 5. A fuller account is provided by Levallois (2017).

Table 5. Key filters used in Gephi and their description

Filter	Use
Inter / Intra Edges	Shows edges which connect nodes with the selected attributes. Several attributes can be selected, meaning edges can be followed within (intra-edges) and between (inter-edges) groups of documents.
Partition	Only shows nodes (and their edges) based on selected attributes. This filter is used to see the documents of a given attribute and their citations. Again, several attributes can be selected.
MASK	Must be used with another filter. It provides additional parameters which allow edges and nodes to be masked based on the direction of edges.

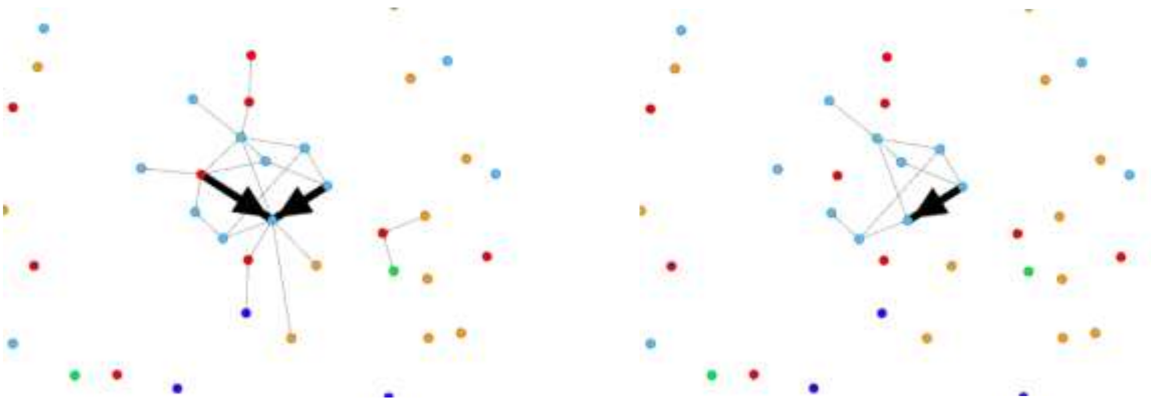

Figure 1. Filtering edges to focus on a subgraph

Checking for a power-law with R

The degree distribution is simply the frequency of nodes of a given degree, where degree is plotted on the x axis and frequency on the y axis. If the plot of the degree distribution provided by Gephi appears to follow a power-law, as shown to the left in Figure 1, an important step involves checking that it truly is a power-law. We used the R programming language (R Core Team 2017). The RStudio IDE is free to download, and the script we used is provided in the Appendix. Simply copy and paste the script into the command window (top left). The first 10 lines install and load the required packages. The next few lines load the Pajek file generated in Gephi. The only change to be made is at line 12, where the directory path in the `file_path` variable must point to your Pajek file. Make sure all backslashes are replaced with forward slashes '/'. As the script is run, results will be printed to the console (bottom left) and graphs will appear on the right.

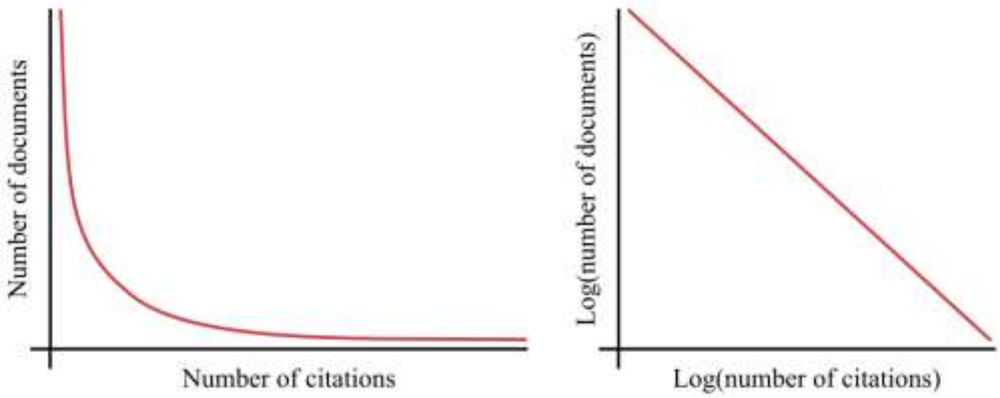

Figure 2. Example of a power-law (left) and the result of linearising the data (right)

To check if the degree distribution (calculated lines 14 to 19) follows a power-law, the literature suggests plotting the degree distribution on logarithmic axes (lines 33-41; Newman 2003: 185-186). If the distribution follows a power-law, scaling each axis logarithmically – what is called linearising the data – should result in a straight downward line, as illustrated to the right in Figure 1. It also helps remove complexity by making non-linear data linear, allowing linear regression to be carried out. The latter allows one to measure how closely the data fits the model (lines 27-31). The call to `summary()` on line 31 prints the results of the linear regression to the console. A high r-squared (close to one) and low p-value (nearly zero) potentially indicate a strong fit with the model. The former indicates how well the model describes the data while the latter represents the probability that this outcome is due to random chance. Linear regression against other models should then be carried out to ensure a power-law is the best fit (lines 43-54 test against an exponential model). The next step is to find the scaling parameter α – the slope of the line – and lower limit for the model – the threshold from which the model applies. This is accomplished with the `fit_power_law()` function (lines 56-60). Finally, they suggest examining the inverse cumulative distribution function (CDF) on logarithmic axes, again to linearise the data (lines 62-90). The resulting plot shows the probability that a node will have a given degree. Since we are looking for a power-law on linearised and inverted data, we should find a straight downward line, indicating the decreasing probability of finding a high degree. The parameters of the model (the scaling parameter and lower limit for the model) are printed to the console at line 81. Again, we check the fit with the model using linear regression, and check r-squared and the p-value through the final call to `summary()` (line 90). All the graphs produced in RStudio can be saved as images by clicking on Export in the plot window.

Bringing it all together – answering the research questions

The nature of the study means the analysis is a recursive process between network and statistical analyses, the results from one step guiding the hypotheses and assumptions being tested in subsequent steps. Answering how evidence is produced only required statistical methods and visual analysis of the network. This involved analysing the data on evidence production and examining the references used by evidence documents, for instance. Answering how evidence circulates and is used required all tools, and rested on a broader analysis of the structure of the network, influential documents, how different areas cite each other, and so on. Table 6 provides some examples of guiding questions we used in our analysis and the appropriate tools to use.

Table 6. Example guiding questions with the corresponding data and tools to use

Research question	Guiding question	What data to use	Analyse dataset	Network metrics	Visual analysis	Analyse degree dist.
Evidence production	What kind of actors produce evidence?	Use the metadata data from the dataset	X			
	To what extent are evidence documents transparent about evidence production?	Use the evidence production data from the dataset	X			
	How do evidence documents use references?	Follow edges between documents and examine document attributes	X		X	
	How do different areas of focus inform each other?	Follow edges between documents and examine document attributes	X		X	
Evidence use	What does the structure of the network imply for its behaviours?	Examine the degree distribution				X
	What are the key documents in the network?	Identify nodes with high degree and betweenness centrality, clusters of nodes, examine document attributes, and so on		X	X	X
	Where are key evidence documents being taken up?	Follow edges between documents and examine document attributes			X	
	What makes influential documents important?	Examine document attributes	X	X	X	

Interpreting power-law behaviour

How the data is treated depends on what structure is found in the degree distribution. We focus on interpreting a power-law below, since this is the most common distribution encountered in citation networks. Finding a degree distribution which follows a power-law has implications for the structure and behaviour of the network. First, the network is structured around key hubs at multiple levels in the graph, meaning any properties and behaviours exhibited at the level of the whole network will be replicated at lower levels. As a result, power-law graphs are also called scale-free networks. Second, network hubs filter redundant signals, thereby increasing how effectively flows travel, while marginal nodes feed novel inputs into the network. Third, the control hubs have over network flows means the network is hierarchically structured. This means there is no 'average' node (in terms of degree). Instead, focus needs to be set on understanding why some nodes are influential and others are not. Fourth, scale-free networks are the result of a process called preferential attachment, whereby nodes entering the network tend to connect to already influential nodes. This effect causes nodes to cluster together and form hubs at multiple scales, which causes network properties to be replicated across levels. Together, these properties mean scale-free networks are efficient structures which transfer information rapidly. See University of Groningen (2019) for a more detailed introduction to power-laws.

In a citation analysis, a scale-free network is structured by key documents which concentrate and redistribute most of the knowledge in the network, and therefore control the 'narrative' and standard practices of an organisation. We used vertex-level metrics and visual analysis to identify these documents and understand why they are heavily cited. Our analysis demonstrated that many documents which structure and guide knowledge and evidence in the IFRC are Federation-wide guidelines, policies and other documents produced by the head office in Geneva, for example. If your data does not follow a power-law, it may follow an exponential distribution (also tested for in the R script) or any other model. Averages may be meaningful in some contexts, in which case focus is set on understanding which documents are representative of the average document, which are outliers, and why this is the case.

Conclusion

These guidelines outlined the method and tools used to carry out a citation analysis of an organisation's document base. The first part presented the gathering and inclusion of documents, the creation of the dataset, and how to use network analysis software. The second part introduced the statistical and network tools used, and introduced RStudio to check for a power-law. The final part discussed the how the tools were used to answer the research questions. However, these guidelines are not exhaustive. We highly recommend readers wishing to do their own citation analysis explore the literature on ANT and network analysis. Much of the mathematics behind the core topics in network analysis is simple and can be understood quickly and easily. The ubiquity of this subject means there is a wealth of free information and courses covering it online. We also recommend exploring ANT, in particular Latour and Woolgar's (1986) *Laboratory Life: The Construction of Scientific Facts* and Latour's (1991) *We Have Never Been Modern*, which cover two key themes in Latour's thought used in our theoretical framework.

Finally, our full analysis included graphs, inspired from Latour's work, which depict the relations between authors, organisations and documents. These graphs provided a view of 'what is held together by whom, and who is held together by what' (Latour et al. 1992: 34) by showing important authors and partners, and their relation to the document base and to each other. While this approach is unconventional, it provided a view of the human network surrounding document production in the IFRC.

As a descriptive exercise which examines the relations between different elements, network mapping is an ideal tool for shedding light on opaque systems. Beyond its use for citation analysis, mapping techniques can be applied to a variety of complex situations and merits a place the humanitarian toolkit. While our methods and practices continue to evolve, a critical look at the evidence we use and the structure of our knowledge-base is still wanting. In opening this black box, we hope to start a critical discussion on how evidence is produced, circulated and used by aid and development organisations and their staff.

References

Bastian M., Heymann S., and M. Jacomy (2009) Gephi: an open source software for exploring and manipulating networks, presented at the International AAAI Conference on Weblogs and Social Media, 17-20 May 2009. Home page: https://gephi.org

Clarke, P. K. and Darcy, J. (2014) Insufficient evidence? The quality and use of evidence in humanitarian action, ALNAP/ODI: London, 87pp

Christoplos, I., Knox-Clarke, P. Cosgrave, J. Bonino, F. and J. Alexander (2017) Strengthening the quality of evidence in humanitarian evaluations. ALNAP Method Note 2017, ALNAP/ODI: London, 40pp

Davey, E., Borton, J. and Foley, M. (2013) A history of the humanitarian system: Western origins and foundations, HPG Working Paper, Overseas Development Institute: London, 60pp

Ham, K. (2013), OpenRefine (version 2.5). httep://openrefine.org. Free, open-source tool for cleaning and transforming data, Journal of the Medical library Association 101(3) 233-234

Hankey, W. and Pictet, G. (2019) Following evidence from production to use at the International Federation of Red Cross and Red Crescent Societies: where does it all go?, Knowledge Management for Development Journal 14(1), 38-66

Kalamaras, D. (2015) Social Network Visualizer (SocNetV). Social network analysis and visualization software. Home page: https://socnetv.org

Latour, B. and Woolgar, S. (1986) Laboratory life: the construction of scientific facts, Princeton University Press: Princeton, NJ, 294pp

Latour, B. (1993) We have never been modern, Harvard University Press: Cambridge, MA, 168pp

Levallois, C. (2017), Using filters. Accessed 20/08/2019 at: https://seinecle.github.io/gephi-tutorials/generated-html/using-filters-en.html

Long, A. (2005) Evaluative tool for mixed method studies. Accessed 15/01/2017 at: https://usir.salford.ac.uk/id/eprint/13070/1/Evaluative_Tool_for_Mixed_Method_Studies.pdf

Newman, M.E.J. (2003) The structure and function of complex networks, SIAM Review 45(2), 167-256

Newman, M.E.J. (2010) Networks: an introduction, Oxford University Press: Oxford, 772pp

Ognyanova, K. (2012) COMM 645 Handout – Gephi basics. Accessed 20/08/2019 at: http://www.kateto.net/wp-content/uploads/2012/12/COMM645%20-%20Gephi%20Handout.pdf

R Core Team (2017) R: a language and environment for statistical computing, R Foundation for Statistical Computing: Vienna, 2673pp

University of Groningen (2019) Scale-free networks. Accessed 09/09/2019 at: https://www.futurelearn.com/courses/complexity-and-uncertainty/0/steps/1855

About the authors

William Hankey is the Director of Black Box Evidence, based near Geneva, Switzerland, which explores how to improve research and its use for social and technical interventions in complex settings. This includes working with development actors to achieve their objectives more efficiently and effectively. Principal contact for correspondence.
Email: will.hankey@bbevidence.com

Gabriel Pictet is Lead for Research, Evidence and Analytics at the International Federation of Red Cross and Red Crescent Societies in Geneva, Switzerland, where he is tasked with strengthening the use of evidence in policy, programming and operations.
Email: gabriel.pictet@ifrc.org

Appendix: Tool R code

```
# Install the libraries (needs internet connection)
packages <- c("igraph", "dplyr")
install.packages(packages)

# Load the libraries
library(igraph)
library(poweRlaw)
library(dplyr)

# Import the Pajek file
file_path <- "E:/IFRC/citationTest.net"
citation_graph <- read_graph(file_path, format = "pajek")

# Get the degree distribution
degree <- degree(citation_graph)
graph_hist <- as.data.frame(table(degree))
graph_hist$degree <- as.numeric(as.character(graph_hist$degree))
# Have a look
graph_hist

# Remove isolated nodes to test for a fit with a powerlaw
graph_pl <- graph_hist %>%
  filter(degree > 0)

# Check the degree distribution for a powerlaw
# First construct the model using the truncated data
logEstimate <- lm(log(Freq) ~ log(degree), data = graph_pl)
logYpred <- predict(logEstimate)

# The summary of the model shows the r-squared and the p-value
summary(logEstimate)

# Plot the data points and line of best fit
plot(Freq ~ degree, data = graph_pl, type = "n", log  = "xy", main = "Log-
log plot")
points(Freq ~ degree, data = graph_pl)
lines(exp(logYpred) ~ degree, data = graph_pl, col = 2)
abline(h = c(seq(1, 9, 2), seq(10, 90, 20), seq(100, 1000, 200)), lty = 3,
col = colors()[440])
abline(v = c(seq(1, 9, 2), seq(10, 90, 20), seq(100, 1000, 200)), lty = 3,
col = colors()[440])

# This time check the fit against an exponential model
expEstimate <- lm(Freq ~ log(degree), data = graph_pl)
summary(expEstimate)
expYpred <- predict(expEstimate)
plot(Freq ~ degree, data = graph_pl, type = "n", log  = "x", main =
"Exponential plot")
points(Freq ~ degree, data = graph_pl)
lines(expYpred ~ degree, data = graph_pl, col = 2)
abline(h = c(seq(1, 9, 2), seq(10, 90, 20), seq(100, 1000, 200)), lty = 3,
col = colors()[440])
abline(v = c(seq(1, 9, 2), seq(10, 90, 20), seq(100, 1000, 200)), lty = 3,
col = colors()[440])

# Pass the object through the power.law.fit() function to find
# alpha, the lower limit, and the KS.p value
graph_dd <- degree_distribution(citation_graph, mode = "total")
plf <- fit_power_law(graph_dd, implementation = "plfit")
head(plf)
```

```
# Calculate the Inverse Cumulative Distribution Function (CDF)
occur = as.vector(table(graph_hist$degree))
occur = occur/sum(occur)
p = occur/sum(occur)
y = rev(cumsum(rev(p)))
x = as.numeric(names(table(graph_hist$degree)))

# Plot the Inverse CDF
plot(x, y, log="xy", type="l", xlab = "Degree", ylab = "Probability",
     main = "Inverse Cumulative Distribution Function")
abline(h = seq(0.2, 1, 0.2), lty = 3, col = colors()[440])
abline(v = c(seq(1, 9, 2), seq(10, 90, 20), seq(100, 1000, 200)), lty = 3,
col = colors()[440])

# Calculate the powerlaw line for the model
pld <- as.numeric(names(table(graph_pl$degree)))
mod <- displ$new(pld)
vals <- estimate_xmin(mod)
# Get the parameters of the model
mod$setXmin(vals$xmin[1]); mod$setPars(vals$pars[1])
# Plot the powerlaw line
lines(mod, col=2)

# Check the fit with the line
cdf_dd <- data.frame(degree = x, Freq = y)
cdf_dd <- cdf_dd %>%
  filter(degree > 0)
logCDF <- lm(log(Freq) ~ log(degree), cdf_dd)
summary(logCDF)
```