

IssueCrawler.net – Scenarios of Use for NGOs and other network researchers (A Generally Useful Document)

Overview

The Govcom.org Foundation, Amsterdam, has developed a software tool that locates and visualizes networks on the Web.

The Issue Crawler, at <http://issuecrawler.net>, is used by NGOs and other researchers to answer questions about specific networks and effective networking more generally. You also may do in-depth research with the software.

The following is an overview of the types of networks commonly sought. For each network type, methods are provided. Finally, there are sample questions, as well as a simple survey to help get you started.

For more information about how to operate the software efficiently, please also consult the instructions of use at http://issuecrawler.net/Issuecrawler_instructions.htm.

Issue Crawler Applications for Civil Society

Common networks sought by NGOs, and the methods used to find them with the Issue Crawler.

1) My Social Network: This is the organization's overall social network. Everybody wants to see their Web network. The network map provides indications about which organizations are in the network (NGOs, media, governments, inter-governmental organizations, donors, corporations, scientific establishments, individuals, etc.). The map also may provide indications of the organization's overall 'centrality' in the network, and/or the 'cluster' it finds itself in.

Method: Does your organization belong to a caucus, a campaign, an association, a partnership, etc.? Is it itself a 'network'? Does it have project partners, or frequent associates? Issuecrawler.net will map the network and provide indications of your organization's showing in it. Use the URLs of all the organizations in a *particular nominal grouping*, e.g., the caucus, campaign, network, association, partnership, or project. Type or paste the group's URLs into the Issue Crawler, and run the harvester.

Crawler Settings:
privilege starting points (on)
analysis 'by page'
iterations of method: 1
crawl depth: 2

Note on crawl depth.

If you have 'links' pages as URLs (i.e., pages with links to the organizations in the group), use one layer deep. If using homepages, use 2 or 3 layers deep. For best results, use links pages. Crawls using homepages consume more crawler energy, and take longer.

Note on crawl name.

Use the group name as your crawl name, e.g., Sixteen Days of Activism Against Gender Violence - Organizers.

2) Their Social Network (An organization's network profile): Use the same method as My Social Network, but from the perspective of another organization.

Crawler Settings:

privilege starting points (on)

analysis 'by page'

iterations of method: 1

crawl depth: 2

3. Issue Network: This is the network of organizations around a particular issue, and the original purpose of the software. Who's doing 'conflict timber'? Who's doing 'communication rights'? What's the network around an issue at this time? Besides organizations, the network may have key documents, events, products, tools, slogans and more that bind the network, or particular clusters in the network. You may explore these commonalities once you have located a network.

Method: Doing a key word search in Google and using all the top ten or twenty results is one way to start, but it is not wholly advisable, for Google's returns rely on the 'entire Web', while we are interested in only parts of the Web – networks. See also pieces that touch on the Issue Crawler philosophy, such as <http://www.chem.uva.nl/easst/sept2002.html>.

To begin locating an issue network, use a short list of URLs which, in your view, provide a good overview of the issue. To gather such a list, you could use Google or another search engine (collecting one or more of the returns), but you also could ask an 'expert', gather organization names or URLs from one or more decent newspaper articles, rely on a particular organization's link list related to an issue, scrape URLs from a discussion list (archive), etc.

Type or paste URLs in the Issue Crawler, and harvest. The Issue Crawler's default settings are for issue network location.

Crawler settings:

Privilege starting points (off).

analysis 'by page'

iterations of method: 2

crawl depth: 2

4. Establishment Network. Most issue professionals understand the current 'establishment' in a particular issue or policy area. To find out what the network location software understands as the establishment, use the strategy for selecting starting points from the Issue Network, and use 3 iterations of method. Be warned, that three iterations of method gives the crawler the most amount of work to do, and the results are slow in coming.

Crawler settings:
Privilege starting points (off).
analysis 'by page'
iterations of method: 3
crawl depth: 2

5. Event Network. This is the favourite of many event organizers and attendees. Who is here? Who should be here? The software locates the extended network of organizations around the event attendees.

Method: Put up a big sheet of paper near the registration desk or common area. Have the attendees write down their URLs on the sheet. If it is an event with many participants, take care in using URLs that are links pages.

Crawler Settings:
privilege starting points (on)
analysis 'by page'
iterations of method: 1
crawl depth: 2

6. Network Evolution. Many NGOs and others are interested in how a particular network evolves over time. Which groups are becoming more central, which less so and why? Is the network shifting geographically? Has it shifted its focus? Once you have located a network, use the Scheduler to re-locate the network at regular intervals. The Scheduler has two settings. Schedule network location according to your original starting points, or according to the last available network. The former is advised, for evolution is more likely to be gradual, but the latter may be more intriguing. (The govcom.org researchers have not explored the latter as of yet.)

Scheduler settings:
every month
original starting points

General insider tips

These tips are written from the perspective of someone who has used the Issue Crawler for some years now.

- 1) Avoid using big media sites, big portals, search engines and similar as starting points for the crawler. All these types of sites link all over the place, and do not produce 'networks'.
- 2) Blogs. If you construe a particular blog as similar to above, avoid it.
- 3) Quantities of starting points. Less is more.
- 4) Find the right links pages, and use them as starting points. A links page is one or more pages on a site that contains hyperlinks to other sites. On a site like Greenpeace.org, there are many links pages. Use the one that pertains specifically to the issue or phenomenon under question. If the links are spread out through the site, or are in a database, try to get a grasp of the site structure. Use the page as a starting point that you believe will lead you to the right links.

- 5) Stripping URLs from sites. If you cannot see the URLs on the page, view source in your browser, and copy and paste the portion of the html code with links to other sites into the Harvester. The Harvester will strip out the code, leaving only URLs.
- 6) Review your starting points before launching a crawl. Have a look at your starting points. Delete duplicates, as well as URLs that are not likely to lead the crawler to links.
- 7) Exclusion list. Make your own stop list. The stop list is a list of sites or pages the Issue Crawler neither will crawl, nor return in a network. If you are searching for networks in Russia, there are site stat counters, big portals, software download pages, and other sites particular to an area (of the Web). Use a stop list for that area. You simply may add your stop list to the default stop list, or paste over the default. Take note of the format of the stop list, and replicate it.
- 8) Keep a log. If you are undertaking a (fairly) serious piece of work, write down your thought process as well as your preliminary discoveries as you go along. These notes will come in handy when you are explaining your results and your interpretation of the network.
- 9) Retrieving your starting points if you did not save them. The starting points are at the bottom of the xml file that contains your crawl results. Copy and paste them into the harvester. Add or delete URLs, and press harvest. Save results for later.

Doing a project with the Issue Crawler.

The Issue Crawler is suitable for the study of networks. Whilst it may be used to get a quick picture, it is not a piece of software like Google, with instant results. Rather, the Crawler takes its time. Leave yourself a few days to a week to find a decent network. The absence of a network is also a 'finding'.

Here are a few questions we have put to the software in the past, during our workshop series, the Social Life of Issues. See also <http://www.issuenetwork.org/> and <http://www.govcom.org/workshops.html>.

Sample questions

1. Networking effects. Has there been a network effect? A network effect may be defined as the uptake of an organizational campaign by an existing network, or the growth of a network around a particular campaign or issue. Examples include the rapid growth of the international campaign to ban landmines, as well as the worldwide protests against the War in Iraq. Examples of intense networking that have yet to yield major social change include the Burma campaigns. Read more at http://www.govcom.org/gco_projects/workshops/6/presentation/.

2. Regional networks. Is there a regional network around an issue? Does 'global civil society' fragment regional civil society? A regional network comprises organizations from a particular region, such as the Caucasus, Central Asia, South Central Europe, or Scandinavia. Whilst the U.S. has its own networks, it has proven difficult to find regional networks without U.S. reliance.

2. Donor effects. Which networks of organizations around issues hold together if donors and/or intergovernmental organizations are removed? Certain actors may understand a donor-free network, or cluster, as more 'authentic'.

3. No Internet. Mapping issues in regions with low connectivity, low Internet penetration. Which issues in countries or regions with low connectivity resonate the most on the Internet? Are these the most relevant issues on the ground? What is the network's understanding of the issues in, say, the Fergana Valley? View case study maps at <http://www.govcom.org/publications/drafts/fergana.html> and [fergana2.html](http://www.govcom.org/publications/drafts/fergana2.html).

4. A virtual society? Interpenetration of the online and the offline. Does the information available online overlap significantly with the offline? For example, are newspaper accounts of what's going on similar to the network's accounts?

5. Network Evolution. HIV-AIDS in Russia. Which organizations have risen (and which fallen) in significance over the past two years, within the HIV-AIDS networks in Russia and Ukraine? What conclusions may be drawn about the type of information on offer from these network dynamics? We found that despite international funding of sex-related HIV-AIDS information agencies, the intravenous-drug-use information providers, in Russia, are still the most significant in the networks. View and print large-format case study map at http://www.govcom.org/publications/drafts/public-health_HIV-AIDS.pdf.

6. Doing without news? Has my initiative (which received press attention) resonated as well in the broader issue network? We may have a press strategy. Do we need a network strategy? Read more at <http://www.issue-network.org>.

7. Does a network have preferred formats? Which formats circulate best in networks? What does a network do with a press release? What does it do with a tool or a prize?

Issue Mapping Process with Survey and Methods

The Issue Mapping process could begin with a survey. The purpose of the survey is to compile specific collections of URLs for inputting into the software, and locating a network. Additionally, the survey aids with particular kinds of analyses, explained in the methods sections below.

Survey

1. Name your issue area.
2. Name the most significant organizations in your issue area, with URLs.
3. Name the most significant sub-issues, terms, slogans, campaigns, individuals, etc.
4. List the most significant documents in your issue area, with URLs.
5. List the most important conferences in your issue area, for the past year, the current year and next year, with URLs.
6. List the organizations in your issue area that you have had email contact with in the past 6-12 months, with URLs.

Survey results and methods

The survey results provide:

1. An issue (question 1), and the starting points for a crawl to locate an issue network (question 2).
2. The substance that may hold together a network (question 3). Each of these terms either alone or in some combination may determine the 'life' of the network - the blood that circulates around the network, keeping it alive. Once a network is located, peruse it. Note whether the organizations in the network use the same language, refer to the same slogans, etc. One may color code or otherwise annotate the map showing which organizations are engaged in which sub-issues. Such an information overlay enriches the interpretative power of the map.
3. Key documents that introduce you to an issue area. These documents may organize a network, or a cluster. You also may find out who links to these documents, providing further indication of key, knowledgeable actors. To find out who links to a page or site on the Web, use the advanced search option of a search engine. For more exhaustive research, use more than one search engine.
4. Events that may bring together significant actors (question 4), with URLs that may provide participant lists. The results pave the way for locating an event network. They also may help you to organize a future event.
5. Email networks. Having a list of organizations (with URLs) opens the door for comparative analysis between the issue network and someone's social network. Note that, in advance of analysis, an individual often will equate his/her email network with the Issue Network. Findings often show a divergence between the key players and whom one knows.

A few key terms

Crawl depth: Selecting a crawl depth of one means the crawler will retrieve the links from the pages located **here**: <http://www.site.com/>**here** A crawl depth of two means the crawler will retrieve the links from the pages located **here**: <http://www.site.com/>**here/here**

Iteration: One instance of analytical method.

Starting points: The URLs you crawl to locate a network.

Prepared by dr. Richard Rogers, govcom.org
Last revision: 30 January 2003