

# **Capturing Sourcing in Articles For Attribution Networks**

Jonathan Morgan

[jonathan.scott.morgan@gmail.com](mailto:jonathan.scott.morgan@gmail.com)

## **Introduction**

The social networks of reporters and editors include many more ties than just the strong, direct ties between news reporters and the sources they quote. Journalists are connected to other types of newsroom employees, to employees in other parts of their news company, to the myriad people reporters and editors live near and interact with, and to the range of relatives and friends of all the people they know personally, some of whom are a number of degrees removed (people who know people who know journalists, etc.).

In this study, I will outline a strategy and a methodology for approximating the characteristics of newspaper reporters' social networks at different points in time, starting with direct ties to sources reflected in news articles. The resulting social network maps will allow researchers to begin to more explicitly quantify and classify the effects of layoffs and consolidation on newsrooms. Multi-level, multi-theoretical social network analysis also offers a way to find theoretical patterns in changes in networks of news resources, patterns that can be translated into more precise theories of how these networks react to internal and external pressures, including downsizing.

## **Data Processing Procedures**

There are no special procedures for preparing content for coding once the dates for coding are selected. We will move through the issues of the newspapers chronologically, issue by issue and page by page. For each article, the coders will:

- 1) Read through the story to identify topics.
- 2) Record the traits of the story and of the reporter(s) of the story.
- 3) Go through the story again looking for sources, organizations, and locations per the operational definitions for each outlined below, and recording information on each as they appear in the story. This information will include the number of times the source is quoted.

## **Content Variable Operational Definitions**

The basic unit of measurement in this study is the article. This study breaks articles out by publication, publication date, newspaper section, and newspaper page. It then examines the author(s) and source(s). In each article, the coder will identify and categorize the following:

- Traits of the article itself, including publication date, section and page number, and primary and secondary topics.
- The reporter(s) or author(s). If an article is a letter to the editor, the author is the person who submitted the letter. If the article is an editorial or an article by "staff", then the reporter is left blank.
- The sources in the story, including organizations and documents.
- Any locations that are a part of the story, not including locations that are simply mentioned in the story.

## **Article**

The basic unit of measurement in this study is the article. For the purposes of this study, an article is considered any locally produced piece of content within any section of one of the selected newspapers. In the news section, this includes letters to the editor and op-ed pieces. In sports or features pages, this would include local letters. No wire service or syndicated articles will be counted as part of this study. Editorials and articles attributed to “staff” are included, but will not be assigned a reporter. They will be included in case they include mentions of sources, but will not be assigned to any reporter.

Coder will identify and categorize the following at the article level:

- *Unique ID* for the article
- The *newspaper* in which the story appeared.
- The *date* of the issue in which the story appeared.
- The *section* in which the story appeared.
- The *page* on which the story appeared.
- The *headline* of the article.
- The *text* (if available and easily captured. If not, forget about it).
- The *main topic* of the story (see more detailed information on topics below).
- *Other topics* of the story, ranked in order of how prominent they are in the story, 1 being most prominent. This will be a catchall, and essentially, we will capture rank data, but we might end up converting these to binary for other topics present or not, since the rankings will have vastly different meanings.
- *Article type*: News, Sports, Feature, Other
- The *reporter(s)/author(s)* of the story (details on this coding are below).
- The *sources* contained within the story (details on this coding are below).
- The *locations* covered in the story (details on this coding are below).

## **Reporter(s)/Author(s)**

The reporter/author (simply called author for the rest of this section) of a given piece of content will be either the person(s) credited in bylines for the article or the person(s) who authored a letter or op-ed. Every distinct author gets a separate entry. For each author, we will code the following:

- The *name of the author*. Should capture as much of the name as appears in the byline of an article, or in the tagline of a letter to the editor, to make it easier to tie sources together across articles.
- *Gender of the author*: male, female, NA. This can be determined by first looking at the gender of pronouns referring to the source (“he” = male, “she” = female), then by trying to identify the gender the first name of the source would normally represent. If gender cannot be ascertained, then enter NA.
- *Author type*: news staff, editorial, government official, business representative, other organization representative, public. If the story is a news or feature story, then consider the author a “news staff” member. If the story is attributed to the newspaper but not to a particular reporter, don’t choose an author. If the story is

an unsigned editorial, choose “editorial”. If you can’t determine the author’s type, just record them as a member of the public.

### **Source(s)**

Sources in a given piece of content will be any person, organization or document specifically cited as a source of information. To be considered a source, the provider must be identified with a verb of attribution, which is a statement of direct or indirect communication. Direct communication in news stories is determined by verbs or attributions such as "said," "reported," "stated," and "noted." Indirect communication included verbs related to mental states such as "hopes," "feels," and "believes." Each distinct source should get its own separate entry. If an organization is cited as a source, it gets its own entry, but if subsequently a spokesman is designated as speaking for that organization, that person becomes the source, and you capture the organization as the organizational affiliation for that person. Documents are considered sources only if they are not presented by someone. If a document is noted as having been provided by a person, then the person is the source, not the document, and the document is not counted as an additional source. For each source, we will code the following:

- *The name of the source.* Should capture as much of the name as appears in the article, to make it easier to tie sources together across articles. If the source is an organization or document, specify the name of the organization or as much of the title of the document as is present in the article.
- *Gender of the source:* male, female, NA. Organizations and documents do not have a gender. For people, this can be determined by first looking at the gender of pronouns referring to the source (“he” = male, “she” = female), then by trying to identify the gender the first name of the source would normally represent. If gender cannot be ascertained, then enter NA.
- *Title:* Title of source, if given.
- *Organizational affiliation:* should capture as much information on the source as is possible. If the name is present, capture the name. If affiliation is only spelled out in general terms, capture the description of the organization, and factor this into selection of source capacity.
- *Source type:* anonymous, individual, organization or document.
- *Source capacity:* government source, business source, other organization source, personal source, expert, other. More details on each:
  - Government source – sources who speak on behalf of the government, or on the behalf of a government agency. Could be from any branch of government (executive, judicial, legislative). If a source is speaking their own opinion, even if their organizational affiliation is presented, then they are considered a personal source, not a government source.
  - Business source – sources who speak on behalf of a business. If a source is speaking their own opinion, even if their organizational affiliation is presented, then they are considered a personal source, not a government source.
  - Other organization source – sources who speak on behalf of some other sort of organization. Could be non-profit organizations, charitable organizations, etc.

- Expert – someone who is portrayed as being an expert on the topic on which they speak.
- Personal source – person speaking on their opinion, not related to their affiliation with an organization.
- *Source use within the article*: Will be a set of the following items, each of which would contain a count - direct quote, indirect quote, from a press release, spoke at actual event, other. Each time the source is used a certain way, add one to the value for that use. If a source is quoted indirectly, then quoted directly, for example, you would add 1 to both “direct quote” and “indirect quote”. A quote that is two sentences but tied together by a “said” verb in the middle counts as one direct quote. Quotes on different topics count as multiple quotes.
- *Localness of a source*: local, local region, state, national, international, or other. A local source is one who is identified as being located in the city in which the paper is located. A local region source is one who is within the circulation area of the paper. A state source is one outside the paper’s region of coverage, but within the state in which the newspaper is located (in this case, Michigan). A national source is someone who is located outside the home state of the newspaper, but still in the United States. An international source is any source identified as being outside the United States. Other is for a source whose location can not be determined.

## **Locations**

For each article capture all of the locations to which the article pertains. For a location to be counted here, it must be substantially a part of the story, or the setting of a substantial part of the story. The mere mention of a location does not merit its inclusion here. For example, a person being from the city of Okemos isn’t reason to include Okemos if the story isn’t about Okemos. In terms of location, it is most important to capture as specific a location as is possible. So, if you have a city, make sure to capture the city name, don’t just say it is in the county in which the city resides. If you have a specific location, an address, for example, as long as you fill in the city, state, etc., you need not also include the city as a separate location. For each location, capture as many of the following as are present in the story:

- *Name*: name of the location. If the location is just specified as one of the fields below (city, county, state), don’t enter a name here, and enter the name in the corresponding field. This field should just be used for named places that are not a city, county, or state. So, a named neighborhood would go here, or an area of a city described by boundaries. If a location is described, capture as much information about it as is in the article.
- *Address*: Street address of the location, if present.
- *City*: The name of the city specified in the article. Include as much information on this as is present.
- *County*: The name of a county specified in an article. Include as much information on the county’s name as is present in the article.
- *State*: The name of a state specified in an article.
- *Zip Code*: Zip code of location, if known.

## Topics

Since there will be a number of article topics encountered in a census of articles gathered for certain time periods, topics bear closer attention here. The goal in this study is to make sure that the articles that cover the same topic are assigned the same topic code. Topics of articles and of sources' contributions to articles will be stored in a central repository that includes the topic name and as precise a description as possible, so coders can look up the topics already used by others. Need to figure out what to do about geography. Perhaps hierarchical topics? I'll have to look into this.

To seed this database, I will implement a random sample of articles from the time periods we are studying and add the topics found in this article sample. In the process of testing this protocol, we will build up the database of topics as we sample articles for use in practice coding and testing. We will also look at the database and brainstorm potential additional topics that are missing, and we will run the list of topics by practicing journalists and journalism researchers to see if we are missing any obvious topics. We will do what we can to have a representative list of topics before we start coding.

There is a certain serendipity inherent in a study that includes letters to the editor and newspapers in general, however. If coders find a topic that is absent from the repository, they will follow the following protocol:

- They will add a new topic to the repository along with as precise a description as possible, so it is there in case another coder happens on the topic.
- They will then send a message to the other coders that explicitly states the new topic value and describes the topic, and that then asks if the topic has already been encountered, and if it has, if the person has coded this topic in the past using another topic value.
- Each coder will reply to the email. If the topic is clear (i.e. not encountered before, and not ambiguous), then it enters the set of available topics. If the topic is not clear, then we will have to get together and agree on a topic value and precise description that we will then add to the topic repository.

## Code Sheet

---

### Article

---

ID: \_\_\_\_\_

Newspaper: \_\_\_\_\_

Date of issue: \_\_\_\_\_

Page of story: \_\_\_\_\_

Main topic: \_\_\_\_\_

Other topics: 1. \_\_\_\_\_

2. \_\_\_\_\_

3. \_\_\_\_\_
4. \_\_\_\_\_
5. \_\_\_\_\_
6. \_\_\_\_\_
7. \_\_\_\_\_

Article type (Circle one) :

News (byline)      Feature      News (no byline)      Letter      Opinion      Other

-----  
**Authors**  
-----

**Author 1**

Name: \_\_\_\_\_

Gender (circle one):      Male      Female      Unknown

Author Type (Circle one):      news staff      editorial staff      government official

business representative      other organization representative      public

**Author 2 (etc.)**

Name: \_\_\_\_\_

Gender (circle one):      Male      Female      Unknown

Author Type (Circle one):      news staff      editorial staff      government official

business representative      other organization representative      public

-----  
**Sources**  
-----

**Source 1**

Name: \_\_\_\_\_

Gender (circle one):      Male      Female      Unknown

Title: \_\_\_\_\_

Organizational Affiliation: \_\_\_\_\_

Source Type (Circle one):      anonymous      individual      organization      document

Source Capacity (Circle one):    government source    business source

                         other organization source    personal source    expert    other

Source Use Within Article:

Direct Quote: \_\_\_\_\_ Indirect Quote: \_\_\_\_\_ From a Press Release: \_\_\_\_\_

Spoke at event: \_\_\_\_\_ Other: \_\_\_\_\_

Localness (Circle one):    local    national    international

**Source 2 (etc.)**

Name: \_\_\_\_\_

Gender (circle one):    Male    Female    Unknown

Title: \_\_\_\_\_

Organizational Affiliation: \_\_\_\_\_

Source Type (Circle one):    anonymous    individual    organization    document

Source Capacity (Circle one):    government source    business source

                         other organization source    personal source    expert    other

Source Use Within Article:

Direct Quote: \_\_\_\_\_ Indirect Quote: \_\_\_\_\_ From a Press Release: \_\_\_\_\_

Spoke at event: \_\_\_\_\_ Other: \_\_\_\_\_

Localness (Circle one):    local    national    international

-----  
**Locations**  
-----

Name: \_\_\_\_\_ City: \_\_\_\_\_

County: \_\_\_\_\_ State: \_\_\_\_\_

Name: \_\_\_\_\_ City: \_\_\_\_\_

County: \_\_\_\_\_ State: \_\_\_\_\_

Name: \_\_\_\_\_ City: \_\_\_\_\_

County: \_\_\_\_\_ State: \_\_\_\_\_