

# Archival Biases and Futures

**Brian Croxall**

[brian.croxall@byu.edu](mailto:brian.croxall@byu.edu)

 <http://orcid.org/0000-0001-5602-6830>

**Rebecca Sutton Koeser**

[rebecca.s.koeser@princeton.edu](mailto:rebecca.s.koeser@princeton.edu)

 <https://orcid.org/0000-0002-8762-8057>

June 2015

<https://belfastgroup.ecds.emory.edu/essays/#archival-bias>

*This essay was peer reviewed by [Geraldine Higgins](#) and [Nathan Suhr-Sytsma](#).*

## Archival Biases

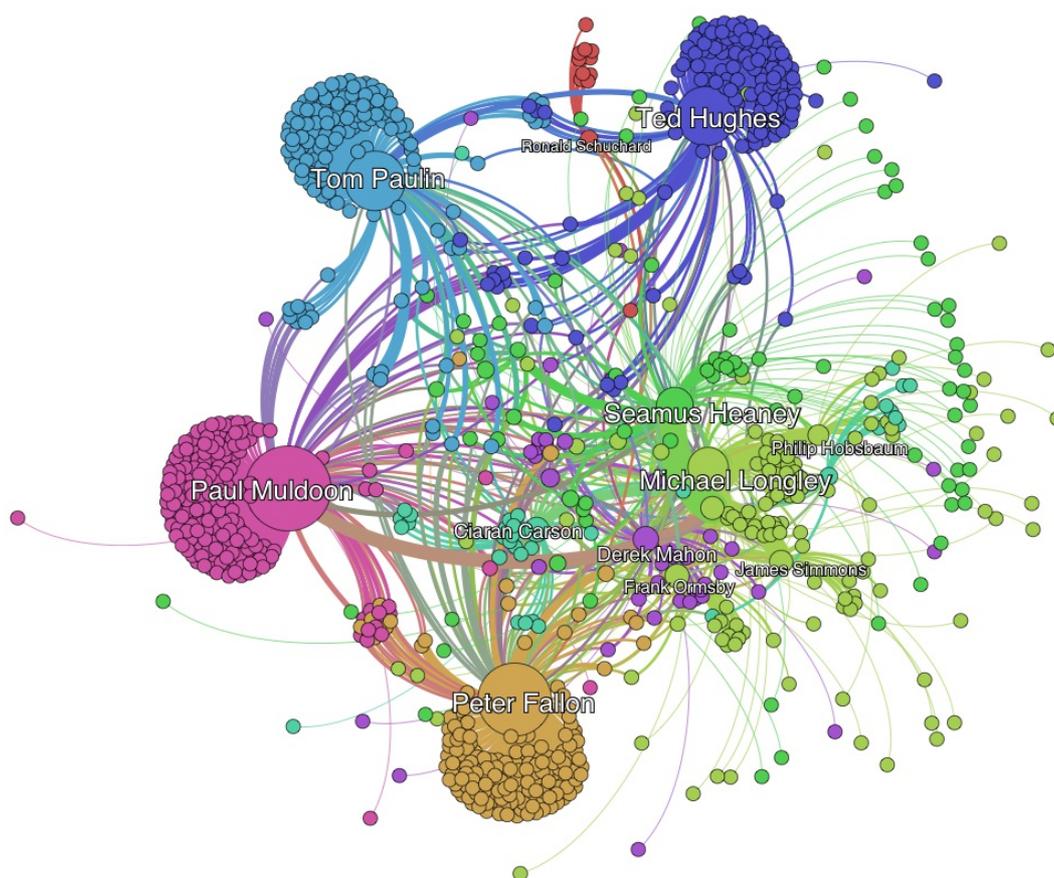
Social network analysis is typically used where data are complete and all connections within a system are known. However, as other humanities networking projects have discovered, building a network based on historical data means that we are inevitably working with incomplete information. In other words, the lack of connections in our graph doesn't mean that no connection exists, but only that we have no documented evidence of one. For a large-scale, historical project like the [Republic of Letters](#), this incomplete information is due to the historic nature of the content they are working with; in other words, not all of the evidence exists any more (see Chang).<sup>[1]</sup> In our case, we have a different bias and different missing data because we have primarily used information from a single archive as the source of our network data.<sup>[2]</sup>

Emory's Manuscript, Archives, Rare Book Library ([MARBL](#)) has a strong collection of Irish literary materials, which is part of what inspired this project in the first place, but certainly there are other archives in Ireland, the U.K., and the U.S. with other materials that would be relevant to this project. For example, Philip Hobsbaum's papers are [held at the University of Glasgow](#), where they are still being processed. Including information about his correspondence would without a doubt increase the number and strength of his connections, as well as introduce other individuals into the orbit of the Belfast Group. As linked open data becomes available in the future, it should be possible to incorporate data from additional archival collections into a project like this.

Even beyond the fact that our data is limited by the selection of collections we've pulled

correspondence information from, our data and network models are biased simply because they are based on data from archival collections. As discussed in [our essay on women in the Belfast Group](#), certain types of people tend to be more prominent in archival collections than others—namely, famous authors. Our network is based on what archives have collected; what archivists have determined is worth keeping; and how much detail those archivists have used to describe those materials. Some manuscripts collections in MARBL are described in much greater detail than others. For example, prior to this project, all the Group sheets in the James Simmons papers were described collectively as “Belfast Group sheets.” To generate more accurate data, we created item-level descriptions of the individual sheets. Similarly, the level of detail in the description of an author’s correspondence varies widely. Our data are incomplete for collections—such as the Derek Mahon papers—where the processing archivists chose to mention correspondents that they considered notable rather than explicitly naming every individual. Observing the gaps in our data is not to impugn the work of our archivists; processing a collection is extremely labor intensive and a complete description of materials would ultimately border on unwieldy facsimile, as the fable from Borges makes clear.<sup>[3]</sup> What’s more, the conditions of a gift to an archive—the terms of which are generally not disclosed—can influence the extent of the descriptions.

The biases inherent in our archival-based data are more evident when looking at the full network generated from our data, especially when algorithmically detected communities are highlighted.



[Download the network graph data \(GEXF\) for this image](#)

We generally see distinct communities for the major figures corresponding to specific MARBL collections, such as Paul Muldoon, Peter Fallon, Tom Paulin, and Ted Hughes; even emeritus Emory faculty member W. Ronald Schuchard has a small community distinct from the rest of the network. The number of distinct communities is most likely the result of the large number of correspondents detailed in each of these collections who do not appear in the other collections. Heaney, Longley, Mahon, and Carson are much more centrally connected in this network representation. While literary historians like Heather Clark would certainly place Heaney and Longley at the center of narratives about the Group, they appear centrally here for another reason: the efforts of our library to collect materials from those connected to the Belfast Group. Our data, again, reflect our collections as much as they reflect reality. When considering the community that the algorithm places the Belfast Group within, one finds several members of the workshop: Longley, Carson, Simmons, Ormsby,

Hobsbaum. This detail surprised us given the fact that, as we noted on in [our previous essay](#), there are no connections between Simmons and Carson on the [network graph of those associated with the Group](#).<sup>[4]</sup> The lack of connection is perhaps explained by Simmons's more frequent participation in the initial gatherings under Hobsbaum, while Carson, who is Simmons's junior by 15 years, could only participate in the second half of the Group.<sup>[5]</sup> Yet, the algorithm placed them in the same community, in large part thanks to their connection to some of the same correspondents. Although this placement is algorithmic, it might also be explained by them both spending almost their entire careers in Northern Ireland, as opposed to Heaney or Muldoon, who worked in the US and other countries. By virtue of remaining in the same environs, Simmons and Carson have a smaller circle of correspondents—at least in our data—which makes it less likely for them to be sorted into another community.

## Archival Futures

The limitations and biases that we've described above are simply the nature of archives, and researchers need to be aware of them when doing any sort of archival work, including data-driven approaches. With these caveats in mind, a network approach to archival data can still reveal new things about literary histories. What's more, a network approach could transform how we use the archive itself.

By applying the same methodology taken for this project, we could enhance data in finding aids from all of MARBL or another large archive. We could then use the resulting data to get a broader sense of the networks within the full archive. MARBL's collection strengths in Irish literature, African American literature and culture, and southern history, would likely appear on a larger scale when we look for communities within the full network of people represented by the archive. But we might also find unexpected connections bridging those communities that would otherwise escape the notice of most researchers or even individual archivists. Visualizing the network of the entire library could be a useful tool for scholars and researchers, enabling them to find related materials in other collections and get a better sense of their research subjects within the context of the archive. Indeed, such an interactive visualization might go a long way to replacing the finding aids as the primary interface to an archive's materials. If multiple institutions began using linked open data in this manner, it would even help a researcher identify other collections in which she or he might look for related materials. Creating such a network could also be a helpful tool for archives to give potential donors a beautiful, big-picture view of the collections in the archive and how their materials might connect.

## Notes

1. ^ Of course it is worth noting that there are gaps in the historical record of events of the Belfast Group, which took place only 50 years ago. For example, while our data set contains information about all the Group sheets that are known to exist, we almost certainly do not know the location of all them. Any of the participants could have kept the sheets from the meetings they attended. We have recently learned, for example, that Arthur Terry kept a couple of Seamus Heaney Group sheets that have been uncollected in any archive (Terry).
2. ^ There are many reasons why our information primarily comes from one archive. The first and most obvious is that since we work at Emory, we have easy access to the documents here. But more importantly, we had the ability to make changes to how the Emory finding aids expose data. We draw directly from the enhanced finding aids to generate the data for the site. By contrast, the finding aids for the Belfast Group materials at Queen's University Belfast are a PDF. While we have incorporated those data, since this project is about enhancing and re-combining library data, working in our own environment is most conducive to demonstrating what is possible. We hope that other libraries will begin exposing their collections in similar ways.
3. ^ Jorge Luis Borges. "Museum: On Exactitude in Science." In *Collected Fictions*. Trans. Andrew Hurley. New York: Viking, 1999. 325.
4. ^ There is a direct connection between Carson and Simmons in our [network graph that visualizes Belfast Group authors by period](#) and shows relationships that we have inferred among Group members based on Group sheet ownership. Carson and Simmons both owned the same Heaney Group sheet. This suggests the following possibilities: 1) they both attended these meetings but didn't know one another or form a relationship that would lead to correspondence or other connections; 2) one or both of them received a mailed copy of the Group sheet but did not attend the meeting in question; or 3) one of them got the Group sheet through some other means. Since the Simmons papers have seven Group sheets from the second period of the Group, it seems likely that he did attend but did not, for whatever reason, end up forming a relationship with Carson that led to their correspondence of such significance that it would be called out in the MARBL Finding Aids. This lack of correspondence, along with their not being described as colleagues or knowing one another in the finding aids, results in Carson and Simmons not being connected in our [visualization of people associated with the writing workshop](#). Of course, we know through others' scholarship and accounts that these two *did* know one another but such information is not present in the description of the collections. Such, again, are the biases and gaps in archival data.
5. ^ In reflecting on the Belfast Group meetings, Simmons says, "I never got to know any of them well, perhaps because I was older" (Dugdale et al. 60). In *The Ulster Renaissance*, however, Heather Clark makes it clear that Simmons did indeed know Heaney, Longley, and Mahon well. Nevertheless he felt increasingly that he was "exclu[d]ed from the poetry community" in Northern Ireland (88; see also 176–182 passim.).

## Works Cited

- Borges, Jorge Luis. "Museum: On Exactitude in Science." In *Collected Fictions*. Trans. Andrew Hurley. New York: Viking, 1999. 325. Print.
- Chang, Daniel et al. "Visualizing the Republic of Letters." 2009. Stanford University. Accessed 4 February 2015. Web.  
[http://web.stanford.edu/group/toolingup/rplviz/papers/Vis\\_RofL\\_2009](http://web.stanford.edu/group/toolingup/rplviz/papers/Vis_RofL_2009).
- Clark, Heather. *The Ulster Renaissance: Poetry in Belfast, 1962-1972*. Oxford: Oxford University Press, 2006. Print.
- Dugdale, Norman et al. "The Belfast Group: A Symposium." *The Honest Ulsterman* 53 (November/December 1976): 53–63. Print.
- Terry, Philip. Email to authors. 29 January 2015.