COMPARING ART HISTORICAL NETWORKS

Doron Goldfarb, Max Arends, Josef Froschauer and Dieter Merkl, Institute of Software Technolog, Vienna University of Technology, Favoritenstrasse 9-11/188, A-1040 Vienna, Austria. E-mail: <vssem@ec.tuwien.ac.at>. URL: <http://vssem.ec.tuwien.ac.at>.

See <www.mitpressjournals.org/toc/leon/46/3> for supplemental files associated with this issue.

Submitted: 22 September 2012

Abstract

This work provides a comparison of link structures present in a common subset of art history related biographic person records/articles from the Getty Union List of Artist Names and English Wikipedia. Throughout recent years, cultural institutions have started to use and contribute to vocabularies like the Getty Union List of Artist Names (ULAN) [1] for more efficient indexing and retrieval. This also includes the explicit encoding of known interpersonal relationships between biographic records of artists and other important art historical actors. Besides such professionally curated resources, the public encyclopedia Wikipedia has recently evolved into a comprehensive knowledgebase, including a large number of articles about art history related persons. Its hypertext form enables the construction of explicit person networks by treating articles as nodes and interarticle links as edges.

The institutional origin of the ULAN vs. its crowd-sourced counterpart raises questions regarding differences or similarities between these two networks. In our work [2], we therefore identify a common subset of person records available in both ULAN and Wikipedia in order to compare the link structures of the two networks. We achieve the identification of matching ULAN / Wikipedia entities by using mappings provided by the Virtual Internet Authority File (VIAF) [3]. This initiative seeks to unify different authority files, including the ULAN, by providing unique identifiers linking records of different authorities that describe the same entities. About 270,000 VIAF records do also contain links to the English Wikipedia. Thus, we are able to identify about 18,000 VIAF entries that contain both ULAN and Wikipedia identifiers. We use the extracted Wikipedia links provided by the DBpedia project for our analysis.

Looking at the data reveals significant differences between the two datasets. First of all, the number of inter-person links is significantly higher in Wikipedia (~54,500) than in ULAN (~9,400). Only 4,465 of the 18,000 persons in the common dataset are connected to at least one person in ULAN, compared to ~16,900 in Wikipedia. Considering Giant Connected Components (GCC), ULAN shows a GCC with ~2,570 persons, incorporating about ~6,400 (~68%) of all the ULAN links, while Wikipedia has a GCC with ~13,500 persons, including ~52,000 (~95%) of all Wikipedia links. Another interesting difference can be found when comparing the temporal distances between interlinked persons. We define the link distance between two person records as the difference in years between their birth dates, being negative for links into the past. Figure 1 shows the distribution of link distances for both link-sets. ULAN links are highly reciprocal, while more Wikipedia links point into the past than into the future. Both datasets have significantly more links with absolute distance of less than 75 years (ULAN: ~99%, Wikipedia: ~89%). This suggests that contemporaries are more likely to be connected than persons from different generations. Wikipedia, however, also contains absolute link distances of more than 2,400 years.

Figure 2 shows a visualization of the Wikipedia links with distances less than 75 years. The nodes are colored according to the respective person’s nationality. The graph has an approximately chronological alignment, starting with the early Renaissance on the left, ending with the present age on the right. Groupings of nodes according to nationalities and historic periods become visible.

The differences between the two link-sets can be attributed to several reasons. The reciprocity of ULAN links and their limitation to distances less than 75 years can be explained by editorial guidelines for their inclusion. The much higher number of links in Wikipedia might be rooted in the freedom of Wikipedia users to include person relationships that would not be considered relevant by art historians. As Figure 2 shows, however, these links don’t seem to be placed at random, but rather reflect the large-scale view of art history through the lens of the English speaking Wikipedia community.

Acknowledgement

This work was funded by the FWF Austrian Science Fund, Project L602.

References and Notes

* This paper was presented as a contributed talk at WebSci2012. See <http://www.websci12.org>.

Figure 1: Distribution of link distances. (© Doron Goldfarb)

Figure 2: Visualization of Wikipedia links < 75 years [4]. (© Doron Goldfarb)