

Reviews in Digital Humanities • Vol. 1, No. 2

Review: Name That Twitter Community!

Melanie Walsh¹

¹Cornell University

Published on: Jan 30, 2020

DOI: [10.21428/3e88f64f.1302b7aa](https://doi.org/10.21428/3e88f64f.1302b7aa)

Project

Name That Twitter Community! (nttc)

Project Developer

[Chris Lindgren](#), Virginia Tech

Project URL

<https://pypi.org/project/nttc/>

Project Reviewer

[Melanie Walsh](#), Cornell University

Project Overview

Chris Lindgren

This Python 3.x module bundles a set of useful code functions for humanistic inquiry of social networks. The module assumes that researchers have a set of network subgraphs created through community-detection,^[1] and they need to more quickly contextualize each community of importance in the corpus for further investigation. It also assumes that researchers have defined periods within their corpus to detect such communities, so they can identify if any detected communities persist across periods.

The module was developed to help researchers answer the following main question: What can these communities be named, and how can they grouped together? This grouping and naming process builds on what Freelon, McIlwain, and Clark refer to as the “hubs” of each community: the top in-degree users from each community and sample of texts that mentions those users (2016, 2018). Yet, extending Freelon et al., this module can also accept each community’s top authors during a period. As a result of this extended hub, the module can also trace potential persistent authorship across communities and generate topic models for each sample to contextualize the hubs over time. Accordingly, the module helps researchers fulfill these contextualizing aims by producing output that answers the following questions:

- What community hubs persist, or are ephemeral, across periods in the corpus, and when?
- Of these community hubs, what are their topics over time?

Overall, the module recognizes the difficulties of coding all of these different dimensions of social network analysis, such as community categorization, naming, and topic modeling. For those reasons, it

aims to help humanities researchers more quickly refine their research questions about particular detected community hubs in a corpus to arrive at more impactful human-centered, yet data-driven, narratives and findings.

Currently, the project has been initially developed and maintained by Dr. Chris Lindgren: an Assistant Professor in the Department of English at Virginia Tech. He coded the module during his own project with Twitter data, which is why the module includes Twitter in the name. He hopes that this review process can help refine the code library, but also broaden its use beyond twitter data to other social network subgraphs, and circulate it among interested research communities across the Digital Humanities.

For more information about the module, its functions, and technical requirements, refer to its project folder on Github.com (<https://github.com/lingeringcode/nttc>) or on the Python Package Index (<https://pypi.org/project/nttc/>).

[1] Numerous community-detection methods exist and can be used, but the use of this module is not contingent on any particular method. However, such decisions matter for the researchers' conclusions drawn from any findings rendered with this module.

Project Review

Melanie Walsh

Project Review

The Python module Name That Twitter Community! (nttc), authored by Chris Lindgren, offers a set of specialized computational tools for social media analysis, a growing area of digital humanities research. It specifically aims to help researchers “name” user communities in a Twitter dataset and follow their participation over time. Researchers who benefit most from Name That Twitter Community! are those with advanced knowledge of Python and Twitter network community-detection, who also have pre-existing Twitter network data ready for analysis.

While reviewing this module, I kept thinking about a question that students who were part of a campus “cultural analytics” group recently asked me. Together we collected millions of tweets related to President Donald Trump’s impeachment, a valuable research accomplishment in itself. But the students were not as satisfied and really wanted to know: “What do we do *now*? What kinds of analyses can we do with the Twitter data *now*?” Detecting, labeling, and tracking salient communities

within this Twitter data through the Name That Twitter Community! module would have been a compelling next step.

I want to recognize the module's broader significance within the field of digital humanities. Like the exclamation mark that concludes its title, Name That Twitter Community! represents an exciting moment for digital humanities social media analysis, as researchers actively build on each other's approaches, theories, and code. Lindgren's module seeks to answer questions central to the work of Deen Freelon, Charlton McIlwain, and Meredith D. Clark, who analyzed #BlackLivesMatter communities in their 2016 report "Beyond the Hashtags: #Ferguson, #Blacklivesmatter, and the Online Struggle for Offline Justice" as well as their article "Quantifying the power and consequences of social media protest." Likewise Lindgren's code follows on the heels of Freelon's [Twitter Subgraph Manipulator](#) (tsm)—a Python module for identifying and analyzing Twitter communities—by helping researchers computationally label the communities that have been identified with tsm or similar approaches. In Lindgren's project one sees the promise of a whole suite of computational approaches to social media data developed by and for the DH community.

The most interesting technical innovation of Name That Twitter Community! is the fusion of social network analysis with topic modeling techniques. The Python module "names" previously detected Twitter communities by topic modeling the text of their tweets, ideally picking up on the main discourses discussed in the tweets. Though Freelon, et al. also used topic modeling in their community detection, they focused on usernames rather than the text of tweets. At the present moment, the nttc Python module is rather difficult to use and requires pre-existing data in specific network file formats (such as [Infomap](#) files). To this end, the project would greatly benefit from clearer documentation and an extended set of examples. These would ideally take the form of downloadable Jupyter notebooks with sample network data that a user could follow. A demonstration of the steps that might be taken to prepare JSON or CSV-formatted Twitter data for analysis with the nttc module would also increase its usability. Despite these limitations, Name That Twitter Community! has great promise for those interested in social media analysis in the digital humanities.