A social network analysis of Twitter: Mapping the digital humanities community

Martin Grandjean

▶ **To cite this version:**

Martin Grandjean. A social network analysis of Twitter: Mapping the digital humanities community. Cogent Arts & Humanities, 2016, 3, pp.1171458. 10.1080/23311983.2016.1171458 . hal-01517493

HAL Id: hal-01517493
https://hal.science/hal-01517493

Submitted on 3 May 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Grandjean, *Cogent Arts & Humanities* (2016), 3: 1171458
http://dx.doi.org/10.1080/23311983.2016.1171458

cogent arts & humanities

DIGITAL HUMANITIES | RESEARCH ARTICLE

A social network analysis of Twitter: Mapping the digital humanities community

Martin Grandjean[1*]

Received: 15 November 2015
Accepted: 02 March 2016
Published: 15 April 2016

*Corresponding author: Martin Grandjean, Department of History, University of Lausanne, Lausanne, Switzerland
E-mail: martin.grandjean@unil.ch

Reviewing editor:
Aaron Mauro, Pennsylvania State University, USA

Additional information is available at the end of the article

Abstract: Defining *digital humanities* might be an endless debate if we stick to the discussion about the boundaries of this concept as an academic "discipline". In an attempt to concretely identify this field and its actors, this paper shows that it is possible to analyse them through Twitter, a social media widely used by this "community of practice". Based on a network analysis of 2,500 users identified as members of this movement, the visualisation of the "who's following who?" graph allows us to highlight the structure of the network's relationships, and identify users whose position is particular. Specifically, we show that linguistic groups are key factors to explain clustering within a network whose characteristics look similar to a *small world*.

Subjects: Internet & Multimedia - Computing & IT; Network Theory; Sociology of Science & Technology

Keywords: digital humanities; social network analysis; Twitter; digital studies; social media; data visualisation; sociometry; networks

1. Introduction: Identifying the *digital humanities* community

At a time where shelves could easily overflow with journal issues and monographs attempting to precisely define the nature of "digital humanities", it seems that we are now, at long last, gradually leaving the "time of definition". Acknowledging that the diversity of these definitions does not help to put an end to the debate, the actors of this field are turning to a more operational concept: the notion of "community of practice". But if this very inclusive concept, just like the "big tent",[1] allows

ABOUT THE AUTHOR

Martin Grandjean is a researcher in intellectual history at the University of Lausanne (Switzerland). He studies the structuration of scientific networks in the interwar period and develops network analysis and visualisation methods for archives and texts. Specialised in data visualisation, he leads parallel experiments in the fields of data-driven journalism, open data and social media analysis. He's member of the board of Humanistica, the French-speaking digital humanities association.

PUBLIC INTEREST STATEMENT

In recent years, the emergence of new technologies in the humanities and social sciences caused a major upheaval. Grouped under the term "digital humanities", thousands of researchers worldwide structure gradually their community around issues related to the use of new tools and methods. Understanding how this new community organises itself is a challenge because it takes very different forms depending on the institutions and scientific disciplines.

This article analyses the presence of the main actors of this community on Twitter, a social media where each user publishes very short messages to his subscribers. By analysing the "who's following who?" network among these 2,500 people, we discover who are the most connected individuals. Language groups are also very visible and allow to question the homogeneity of this community of practice.

© 2016 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

us to overcome the disciplinary clashes, it makes it difficult to identify the borders, as the common denominator seems to be the "kinda the intersection of ..." definition (Terras, 2010). Our study voluntarily chooses to focus on a particular field of expression of this community, a social network that has for many years been regarded as one of the main exchange places for digital humanities. Our goal is therefore not to draw conclusions that go beyond this very specific object, but to observe it in order to offer a transversal view of this movement, otherwise difficult to map with traditional methods. Let's seize the opportunity to ditch the useful "network" metaphor to apprehend it more formally, through a social media that embodies these relationships.

2. Twitter, a growing field of study

Twitter,[2] a social network created in 2006, is a place dedicated to personal expression that brings together hundreds of millions of users around its minimalist concept of microblogging. Its messages of 140 characters and its principle of "following" users without mandatory reciprocity, coupled to a very open application programming interface (API), make it an ideal medium for the study of online behaviour. Its simplicity makes it a frequently used tool to report current events. Hence, many studies analysing the diffusion of information consecutive to an event: an earthquake (Sakaki, Okazaki, & Matsuo, 2010), demonstrations such as the London riots (Beguerisse-Diaz, Garduno-Hernandez, Vangelov, Yaliraki, & Barahona, 2014; Casilli & Tubaro, 2012), international conferences (Grandjean & Rochat, 2014; Jussila, Huhtamaki, Henttonen, Karkkainen, & Still, 2014), teachings (Stepanyan, Borau, & Ullrich, 2010) or interactions on neutral corpus (Darmon, Omodei, & Garland, 2015). These "dynamic" analyses, which typically map networks of tweets, mentions and retweets out, owe their popularity to the availability of the material and the possibility for researchers to analyse its contents. They frequently lead to questions on influence measuring (Subbian & Melville, 2011; Suh, Hong, Piroll, & Chi, 2010), especially when it comes to political communication (Stieglitz & Dang-Xuan, 2012; Vainikka & Huhtamäki, 2015) or scientometry (Haustein, Peters, Sugimoto, Thelwall, & Larivière, 2014).

But when it becomes clear that the content of a user's tweets is not always indicative of his field of specialisation—due to the noise produced by the many personal messages, jokes, politics, etc.—we need to turn to a network whose structure seems more readily analysable in terms of "community": the *follow graph* (Myers, Sharma, Gupta, & Lin, 2014).

3. Dataset

The prerequisite to this study was the preparation of a list of more than 2,500 Twitter users identified as part of the digital humanities community. We saw above that the definition of this field was subject to many changes: rather than stick to lists of association members or authors of a set of journals, we listed all users who identify themselves as being directly or indirectly part of this "community of practice". It is in the very short Twitter "bio" (160 characters) that we spotted the vocabulary linking these researchers together. First of all, it is by listing all the followers of the most visible users (national or international institutions, established professors and researchers in the field, Twitter accounts of scientific events, etc.) and by reviewing their biographies that a first selection was made. Within this corpus, we then randomly select a number of users and we also analyse their subscribers. This list is then enriched in three ways: by the identification of users who tweeted with specific DH conferences hashtags; through the self-reporting of users who, following the publication of blog posts about this research, announced to be part of the corpus; and finally, through harvesting the results of the Twitter search engine on a selection of keywords related to the digital humanities.

By its nature, this corpus cannot aim to be comprehensive, but it should be noted that it offers, unlike most official lists, to include a segment of the academic population (generally non-institutional) who doesn't publish or doesn't normally participate in official events. They did not wait to receive the DH "label" to assign themselves and see themselves as members of this community.

Specifically, this article analyses the "who's following who?" relationships inside a Twitter list containing exactly 2,538 Twitter accounts of individuals or institutions (on 1 October 2015). This network

is obtained after downloading—via the Twitter API—the list of all the followers of each of the accounts, then filtered according to whether they are themselves members of the list or not. It therefore only concerns the relations within this corpus, not the tens of thousands of non-DH users who follow these 2,500 accounts.

4. Result: an apparent small world

At first, the network of digital humanities on Twitter is a form of small world (Milgram, 1967), at least that's what suggests its visual representation[3] (Figure 1). It indeed shows an extremely dense network. Only one cluster seems to detach itself slightly, while another one, nearby somewhat distorts the very circular structure of the network. The size of the circles/vertices is proportional to the centrality degree of the users (the number of connections, followers and followings together), we note that only 11 of them exceed 1,000 connections. In addition, the colour of the circles shows the in-degree (*inbound degree*, their followers only), allowing us to see that only 17 people (white circles) are followed by more than one-third of the users in the corpus. Median user follows 59 Twitter accounts from the list and is followed by 39 of them.

Are digital humanities—whereas describing themselves as a transversal field—finally a closed world where everybody knows everybody?[4]

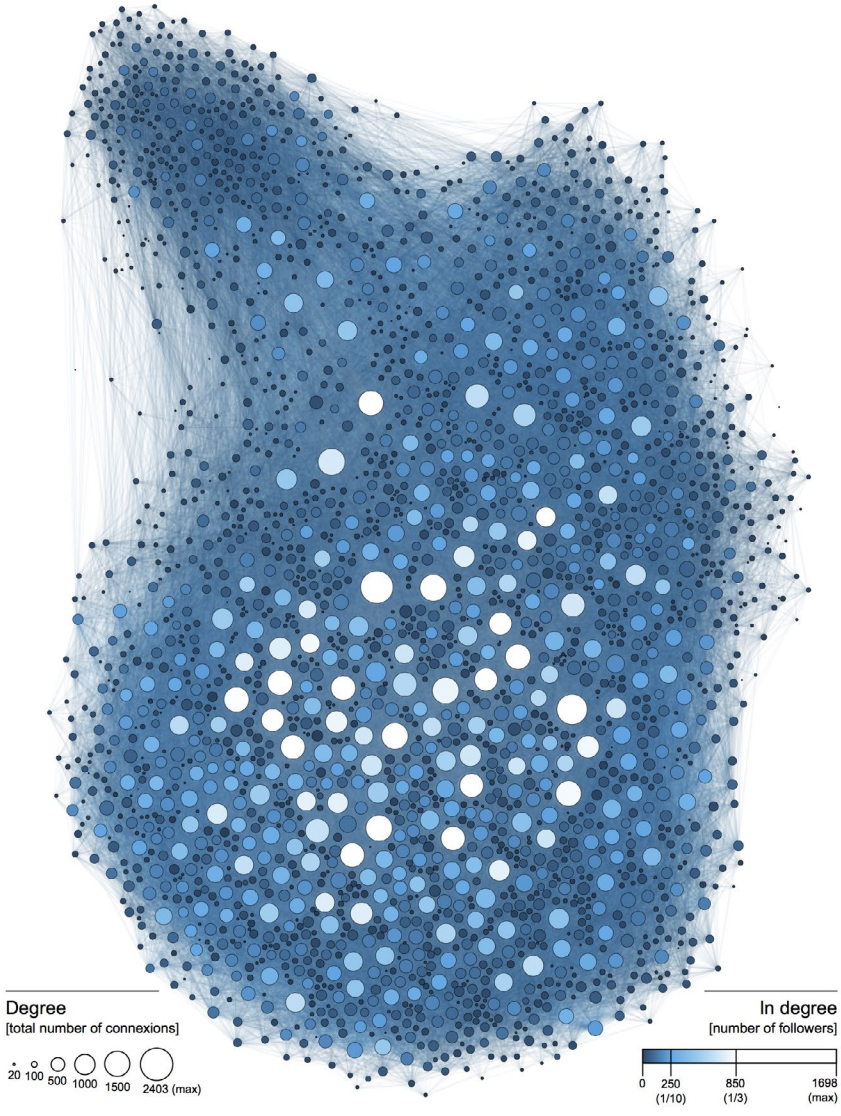


Figure 1. Digital humanities network on Twitter: 2,500 users following each other.

In fact, despite its apparent homogeneity—its limited division into small communities—the density of the graph isn't extremely high. The density is calculated based on the number of possible edges in the network, its value here is 0.036, on a scale from 0 (no edge) to 1 (an edge between every 2,500 vertices).

Even if the network can be structurally considered a small world under the terms of (Watts & Strogatz, 1998), with a high average clustering coefficient (0.366) and a reduced average path length (2.297, with a maximal distance of 5), the application of this concept to an asymmetrical social media remains unclear.

These first elements should not make us forget that this network is a visual representation of a set of data whose complexity is not limited to a simple graphical rendering. Beyond a certain aesthetic, sometimes very suggestive, it is in its ability to generate new research questions—pushing the researcher to get back into the data itself—that a network analysis proves his interest.[5]

5. To follow or to be followed?

Prior to the benefit from more advanced structural measures, the first analysis that we propose is the comparison of the ratio between followers and followings. Figure 2 visualises this relationship as a scatter plot, supported by two bar charts that summarise the distribution of these two values. First observation: more than half of the users follow less than 100 people and are themselves followed by less than 100 people (category A, 63.1%). The vast majority of the corpus is actually made up of very weakly connected users, information that the network visualisation (Figure 1), with its totalising aim, tends to make us forget.

Traditionally, it is considered that users that are highly followed are personalities and institutions whose influence and reputation is superior to users who subscribe to a large number of accounts without themselves being widely followed. We can now distinguish six categories of users, based on their followings/followers ratio (assuming category A users are excluded from this ranking due to their insignificant number of connections):

- Category B: users who follow at least four times more users than they have subscribers (1.3%).
- Category C: users who follow at least twice as many users than they have subscribers (6.6%).

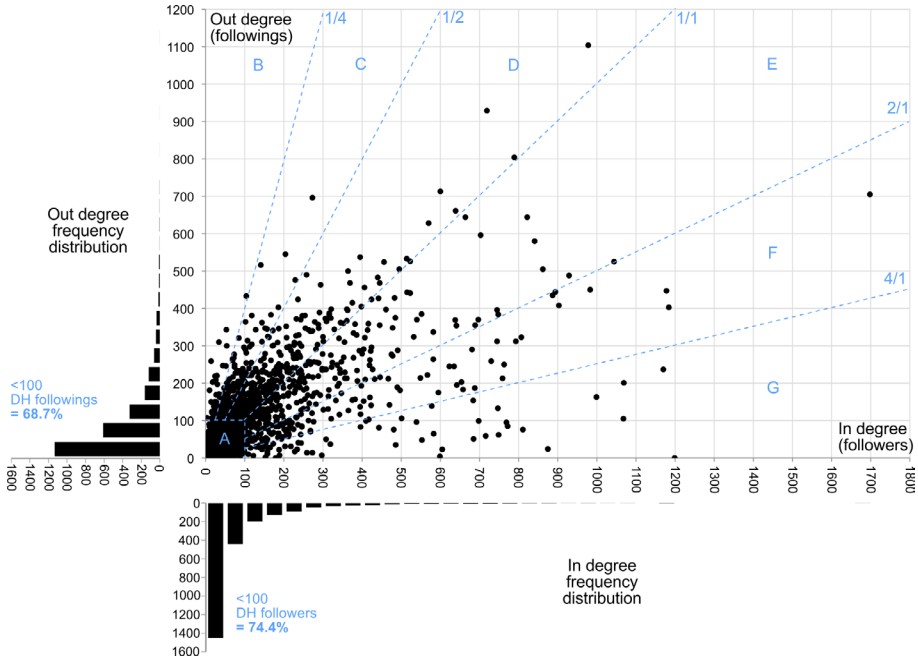


Figure 2. Followings and followers among peers, and frequency distribution.

- Category D: users who follow up to two times more users they have subscribers (13.8%)—This is the largest population of this corpus, behind category A.

The first three categories bring together users who use Twitter a technological monitor. Without necessarily creating content that will make them influencers (even if it's not incompatible), a significant portion of these users is kept informed of the news of their research fields through this social media. It is also to be noted that following a large number of users obviously has a social function that has nothing pejorative. Subscribing to a large number of users typically increases the number of followers (the people being followed are notified of the subscription, they discover their new subscriber and sometimes follow him or her back if interested).

- Category E: users who are followed up to two times more than they follow themselves (8.7%).
- Category F: users who are followed at least twice as much as they follow themselves (4.1%).
- Category G: users who are followed at least four times more than they follow themselves (2.3%).

In the last three categories, we find users who are followed by more users than they follow themselves, generally because they occupy a privileged position in the field (journals, institutions, associations, advanced academic positions, prominent figures in the community or content producers). While the border between categories D and E isn't very significant, the presence of a user in categories F and G is very indicative about his behaviour on the social network. It is indeed among these last two categories that we can find some of the "stars" of the field (in the sense of Moreno, 1934, which lays the basis of network analysis, where the stars are individuals who focus incoming relations).

However, with a little distance, it should be noted that the presence in one or the other of these categories is not a definitive marker of the user's position in the field: having a very high ratio does not always mean being an influential person, but sometimes simply shows a rather elitist attitude (following hardly anyone, e.g.), or a popularity due to an external factor (being a renowned institution outside as well as inside the DH field, e.g.). Let's also recall that we're only analysing the followers/following ratio inside our corpus. A user with a very low ratio may well be followed by tens of thousands of Twitter users outside the community (and a "star" user can have no followers outside this network).

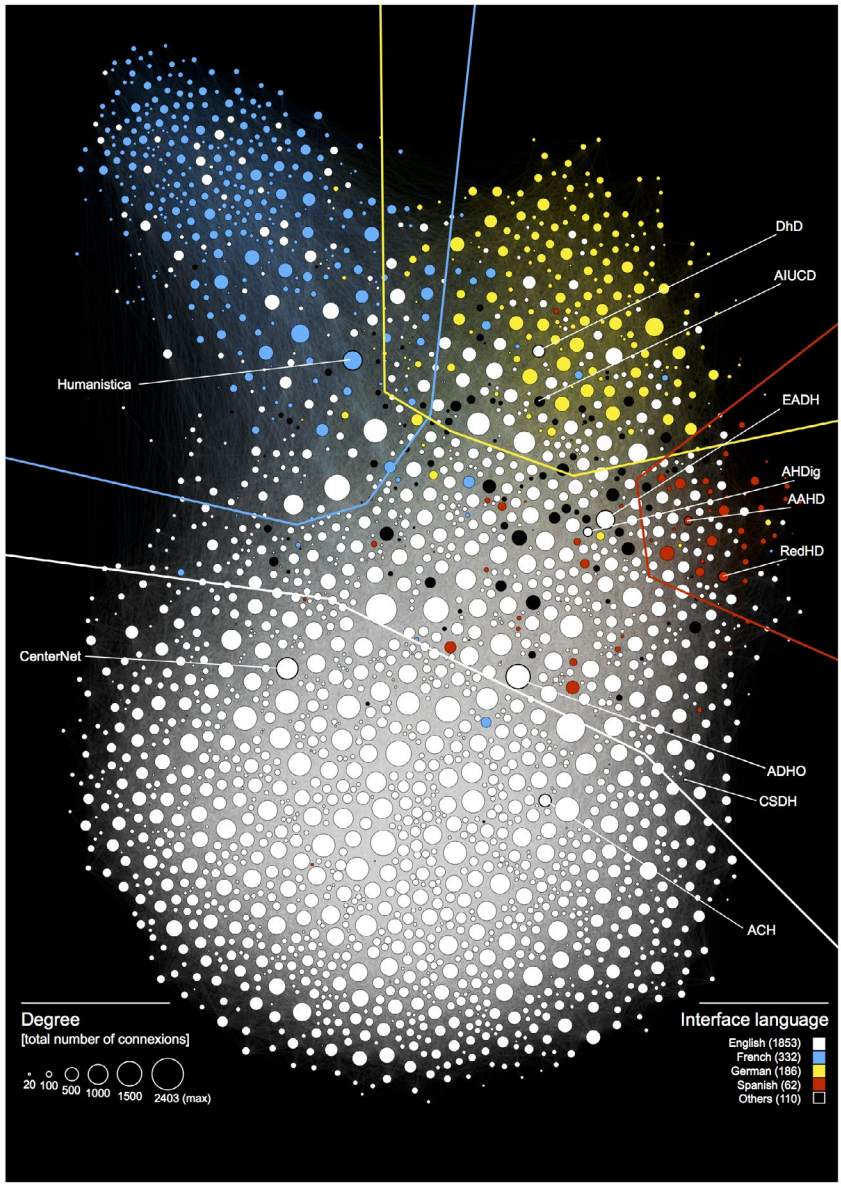
6. A geography of the linguistic communities

Beyond the apparent homogeneity of the network of these 2,500 Twitter users, the geographical, cultural and language distribution must be questioned. While digital humanities are often seen as an essentially English-speaking movement, many local or linguistic communities have emerged in recent years, claiming for their specificities not to be embedded in a large English-speaking congregation. While the geographical issues do not always tally with the language issues (French is spoken in Europe, Africa and North America, Spanish and Portuguese in Europe and South America and English on every continent, at least as a second tongue), national, regional or linguistic associations are emerging,[6] as a "special interest group" of the Alliance of Digital Humanities Association (ADHO), dedicated to the promotion of diversity.[7]

However, the Internet, in general, and Twitter, in particular, are highly globalised places. It is not uncommon for a user to overlook national and linguistic borders as he or she follows the publications of a very wide variety of users. Therefore, are the language communities discernible in our data-set? And if so, how to judge their representativeness regarding the "real world"?[8]

Analysing the language of the tweets posted by users from our corpus for a given period is a chimeric operation, both by the amount of "noise" to disambiguate and by the nature of the content of tweets that are often multilingual. Fortunately, the Twitter API provides, for each of its users, the language of the interface used. Even if English is often the default language, the proportion of accounts using another language is important in our list (27%). Figure 3 visualises, on the same graph

Figure 3. Highlighting the interface language.



as Figure 1, the interface language of our 2,500 Twitter users. It appears very clearly that the two major "clusters" we were already able to distinguish beforehand correspond to very well-defined linguistic communities. In particular, the French-speaking community is almost completely detached from the main group. To a lesser extent, the German-speaking community is clearly circumscribed. Far behind, constituting the third largest non-English speaking community, Spanish-speaking users are also all in the same area of the graph but do not come off from the main group as clearly as the previous two. The remaining users, particularly small Italian-speaking and Dutch-speaking communities, are spread in a kind of "global village" at the intersection of all the other communities.

Two important notes for reading this graph:

- The community of a given language is not limited to the individuals who use Twitter in the concerned language: there are many French or Germans using Twitter in English in the identified

clusters. We estimate that we can add about 30% to the total of the linguistic communities presented in Figure 3, diminishing the English-speaking community in the same proportion.

- The spatialisation of the network is obtained by a force algorithm,[9] which means that the proximity between two vertices cannot be interpreted as a real proximity to each other: as in all network visualisations, this geography is the result of a complex calculation that takes into account each of the edges (there are 236,000).

One thing remains: the French-speaking community is particularly isolated. We can elaborate several hypotheses and questions: is English less used there than in other non-English speaking communities? Or at the opposite, is it a language less mastered in the other regions, justifying that French users are followed less because they are less understandable? Is the French-speaking digital humanities community important and structured enough to be less dependent on English references? Or are the practices so different that the need for skills transfer is less strong with this community than with others? Is it finally only a bias related to the social media analysed, where behaviours differ according to local "cultures"?

Besides, we also note that in the French-speaking cluster, we find most of the French users in the peripheral group. Most of the Swiss, Belgian and Canadian users are rather positioned at the intersection with the other linguistic communities, and thus less isolated.

If the position of the French-speaking community is surprising, it is also because of the comparison with other language communities that we would have expected to be a stronger presence. Rather than seeing the French position as abnormal, is it not worrying to observe such a fusion between the Spanish- and German-speaking communities and the main group? And what about the users using an Italian (36), Dutch (24), Portuguese (10) or other marginal languages (40) interface?

Note that the language distribution within the digital humanities community on Twitter is not comparable with the general distribution of languages in the world, or with the distribution of languages in usually studied tweets sets (Hale, 2014; working on a 2011 data-set). This is not the consequence of a biased data-set but simply a research field that is not (yet) globalised and remains in its major part a European and North American phenomenon (which is also demonstrated by the geography of THATCamps, the emblematic manifestations of this "community of practice", see Grandjean, 2015b).

7. Measuring structural features

Using a formal network only to be satisfied by a comment on its visual characteristics is to miss its structural characteristics. Centrality is a way to quantify the importance of the vertices in a network: its different declinations are frequently used in social network analysis to identify and highlight specific positions (Newman, 2010).

We will therefore seek to go beyond the visual representation in order to list the users of our corpus holding a remarkable structural position. This process is not restricted to online social networking and has been used since the works of Freeman (1978). As in Rochat (2014), Table 1 shows the values of four centrality measurements from our corpus. Two of them, the *In*- and *Out*-degrees have already been exploited above. These are also the easiest to define as they are immediately translatable into Twitter's language, respectively "followers" and "followings". The *Betweenness centrality*, which measures the number of times a vertice is present on the shortest path between two other vertices, highlights users who are structurally in a "bridge" position between the subdivisions of the network. The *Eigenvector centrality* assigns each vertice a score of authority that is based on the score of the vertices with which it is connected. Table 1 is completed by Figure 4, allowing readers to get a sense of the geography of the measurements obtained and to clarify their distribution.

Table 1. Centrality measures of the 100 most followed accounts

User	In-degree		Out-degree		Betweenness		Eigenvector	
	Value	Rank	Value	Rank	Value	Rank	Value	Rank
dhnow	1,698	1	705	5	389,710	1	1.00	1
NEH_ODH	1,199	2	0	2,495	0	2,495	0.81	8
nowviskie	1,184	3	403	45	49,740	27	0.94	2
melissaterras	1,178	4	447	32	92,664	9	0.87	4
dancohen	1,170	5	237	212	35,909	38	0.90	3
mkirschenbaum	1,069	6	201	302	14,958	99	0.86	5
DHQuarterly	1,068	7	105	758	12,813	119	0.72	17
miriamkp	1,044	8	525	17	75,883	15	0.83	7
thatcamp	999	9	163	426	55,051	24	0.68	22
ADHOrg	983	10	450	31	62,772	22	0.68	24
DHInstitute	978	11	1,104	1	268,691	3	0.68	23
GrandjeanMartin	929	12	488	24	313,162	2	0.47	90
HASTAC	903	13	408	43	78,561	14	0.67	28
brettbobley	894	14	444	33	41,450	32	0.83	6
briancroxall	887	15	435	36	35,795	40	0.79	9
kingsdh	875	16	24	1,943	3,503	398	0.51	67
DHanswers	862	17	505	20	63,961	21	0.70	20
trevormunoz	841	18	580	12	58,264	23	0.77	10
elotroalex	822	19	644	8	107,208	7	0.73	12
dpla	811	20	76	1,046	13,987	109	0.58	49
RayS6	807	21	323	105	29,224	44	0.73	14
lisaspiro	793	22	312	112	22,752	61	0.75	11
DHandLib	789	23	804	3	173,640	5	0.58	47
UCLDH	772	24	85	938	9,150	181	0.53	60
kfitz	769	25	95	839	3,391	407	0.71	18
mkgold	763	26	250	187	14,019	108	0.72	16
UMD_MITH	759	27	213	269	12,041	127	0.66	31
chnm	749	28	62	1,226	3,961	358	0.60	41
ryancordell	748	29	384	55	26,960	51	0.72	15
ProfHacker	747	30	133	564	9,073	182	0.59	44
DHcenterNet	745	32	396	50	29,122	45	0.65	32
Ajprescott	745	31	312	113	40,501	34	0.60	40
foundhistory	730	33	259	172	20,907	67	0.70	21
frederickaplan	719	34	929	2	248,304	4	0.50	70
manovich	716	35	59	1,263	6,526	241	0.57	50
DayofDH	703	36	596	11	88,388	11	0.54	57
sramsay	698	37	99	799	3,074	431	0.68	26
Ted_Underwood	697	38	370	65	35,880	39	0.67	29
fraistat	689	39	355	79	19,653	70	0.73	13
sgsinclair	688	40	187	346	11,714	129	0.67	30
julia_flanders	685	41	51	1,391	2,060	570	0.65	33
alanyliu	684	42	154	457	10,640	152	0.62	37

(*Continued*)

Table 1. (*Continued*)

User	In-degree		Out-degree		Betweenness		Eigenvector	
	Value	Rank	Value	Rank	Value	Rank	Value	Rank
4Hum	681	43	290	135	23,971	56	0.57	51
JenServenti	664	44	644	9	53,033	25	0.71	19
eadh_org	658	45	183	357	14,772	100	0.47	89
amandafrench	654	46	203	296	10,916	143	0.68	25
unsworth	642	47	195	320	10,768	150	0.67	27
TEIconsortium	641	48	354	81	67,328	19	0.42	112
dhgermany	639	49	661	7	118,555	6	0.37	147
Jessifer	638	50	369	68	38,218	37	0.53	59
samplereality	635	51	245	196	8,575	191	0.64	34
tjowens	623	52	245	197	13,600	113	0.60	42
DHCommons	605	53	23	1,972	605	995	0.47	91
williamjturkel	600	54	713	4	80,180	13	0.63	35
britishlibrary	599	55	4	2,434	620	986	0.30	196
DSHjournal	595	56	175	380	14,958	98	0.49	75
mljockers	583	57	65	1,188	1,711	631	0.56	52
karikraus	582	58	264	167	11,431	133	0.61	38
adelinekoh	582	59	337	97	22,458	62	0.50	72
GeoffRockwell	578	60	139	526	5,241	294	0.59	46
christof77	570	61	628	10	87,611	12	0.44	102
jenterysayers	567	62	222	253	7,079	223	0.60	43
zotero	553	63	48	1,444	4,065	351	0.45	95
cunydhi	552	64	385	54	30,468	43	0.51	68
scott_bot	549	65	214	268	15,291	94	0.53	61
ernestopriego	540	66	370	66	28,079	48	0.54	56
mia_out	539	68	96	826	6,387	244	0.48	81
omeka	539	67	96	825	5,632	282	0.44	101
jasonrhody	532	69	324	103	9,256	177	0.63	36
clioweb	523	70	441	35	28,267	46	0.60	39
jenguiliano	523	71	526	16	26,969	50	0.59	45
GCDH	515	72	443	34	72,751	16	0.34	167
wragge	514	73	533	15	65,783	20	0.54	55
CathyNDavidson	501	74	106	749	4,624	326	0.47	87
Literature_Geek	497	75	181	368	5,893	270	0.52	66
DH2014Lausanne	495	76	505	21	68,466	17	0.38	140
lornamhughes	491	77	288	137	15,165	95	0.47	88
sleonchnm	490	78	189	341	5,788	273	0.54	54
DARIAHeu	486	79	35	1,695	3,828	367	0.27	228
tmcphers	485	80	375	61	24,907	55	0.54	58
HumanisticaDH	483	81	428	38	103,554	8	0.27	243
DH_Oxford	481	82	76	1,047	3,768	374	0.32	179
GeorgeOnline	478	84	348	83	12,967	118	0.55	53

(*Continued*)

Table 1. (*Continued*)

User	In-degree		Out-degree		Betweenness		Eigenvector	
	Value	Rank	Value	Rank	Value	Rank	Value	Rank
HASTACscholars	478	83	273	155	10,337	157	0.48	85
pannapacker	475	85	277	148	13,316	115	0.52	63
tanyaclement	470	86	142	506	1,777	617	0.58	48
jean_bauer	467	87	212	276	5,368	287	0.52	62
inactinique	461	88	397	49	90,683	10	0.33	171
roopikarisam	456	89	524	18	41,926	30	0.47	86
jamescummings	446	90	216	265	10,870	146	0.41	122
jasonaboyd	445	92	468	27	18,187	75	0.50	71
FrostDavis	445	91	355	80	13,818	110	0.49	78
UCLA_DH	443	93	81	985	4,716	320	0.34	169
DARIAHde	442	94	427	39	67,421	18	0.22	303
Adam_Crymble	440	95	483	25	48,675	29	0.44	103
amyeetx	432	96	211	277	4,040	353	0.52	64
patrick_mj	427	97	237	213	6,861	229	0.50	69
seth_denbo	424	98	296	130	8,269	201	0.49	73
j_w_baker	424	99	424	40	22,959	59	0.40	128
epierazzo	423	100	141	512	5,785	274	0.35	160

7.1. In-degree

The number of followers decreases very rapidly within the top 100 users. The most followed account is @dhnow (Digital Humanities Now[10]) whose aggregation mission seems to be recognised by the community. Then, follow leading figures, institutions, associations and publishers. Spatially, a high inbound degree is not the exclusive privilege of one of the "linguistic communities" studied above. Note that the most followed users are still generally—and logically—located in the "heart" of the network, between the central "global village" and the English-speaking region.

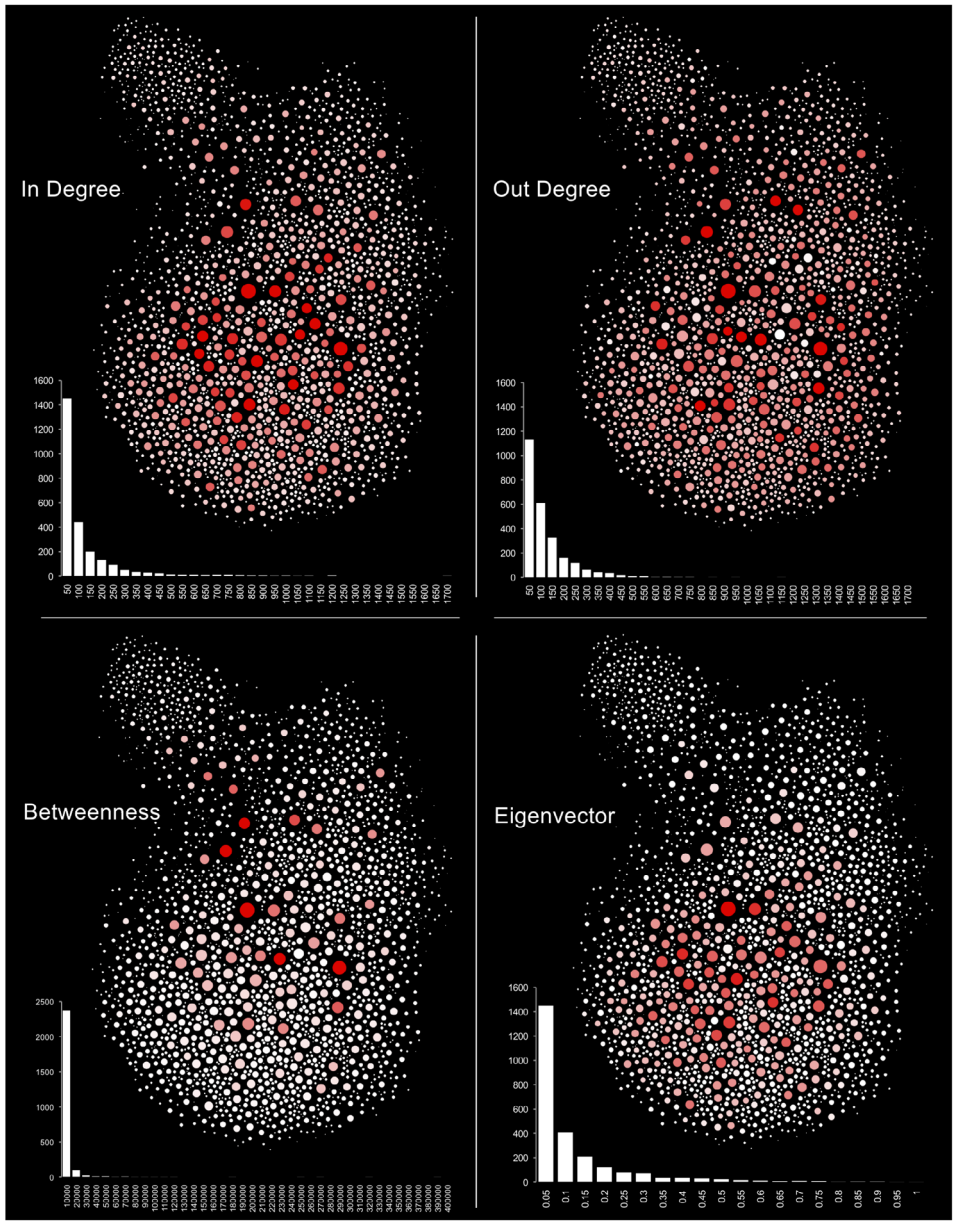
7.2. Out-degree

Some massively followed accounts are themselves following very few users from the list. Consequence: the classification according to the out-degree is quite different from the previous one. The account that follows the largest number of users is @DHInstitute (Digital Humanities Summer Institute,[11] University of Victoria), an event that presumably seeks to bring together the community. The distribution in "long tail" of this measure is less pronounced than for the in-degree. This can be explained very naturally because on Twitter, the majority of users follow more people than they are followed themselves. Except for a few users with a high in-degree but a very low out-degree, the distribution of this measure on the graph is very similar to the previous one.

7.3. Betweenness

In a harmoniously distributed network, highly connected accounts (with a high centrality degree) are usually also the ones most often being on the shortest path between the vertices of the graph. But we have seen that our network contains clusters that are detached from the main structure. It is therefore logical that we find individuals with high betweenness in the area which is located at the intersection between the main network and the French and German clusters. These users—often French or German speakers engaged in international structures as ADHO or EADH—are transmission belts between different regions of the graph.

Figure 4. Spatial and statistical distribution of the four metrics.



7.4. Eigenvector

As the eigenvector centrality is assigned to the vertices according to the score their neighbours received, it produces a result that highlights the very connected users within the larger group of our graph. Here, this measure of authority no longer focuses our attention to the periphery and to the inter-community "bridges", but rather to the centre and its English-speaking majority. Except a few hyper-connected users who monopolise the top positions in almost every centrality ranking, we see here less cosmopolitan users, better "installed" in their English-speaking environment.

We will avoid considering these measures as indicators of influence. They document the network structure, not the nature and content of the relations themselves. They nevertheless allow to pinpoint patterns whose study should be coupled with an analysis of the position of these users in the world of academic hierarchies, publications or co-directions of research projects.

8. Limits and perspectives

The mode of creation of the data-set can, at least partially be a factor in the small-world visual impression: if the majority of the 2,500 users were detected because they were following a more "visible" account, then the high density of the network is logical, even though an effort was made to focus on minorities. Self-determination also has its limits: a person that all his colleagues would describe as a "digital humanist" but describes himself on Twitter with a biography that does not describe his scientific activity may pass under the spectrum of our analysis. Note that it is possible to overcome this problem by no longer focusing on biographies of registered users but on their structural characteristics themselves. The next step of this analysis could indeed be to list the hundreds of thousands of followers "off-list" from our 2,500 selected profiles and automatically integrate to the list those who follow or are followed by a determined number of members from the original list. A systematic way of grouping the community of those who, even without being practitioners, "follow" the latest research in DH.

Concerning the debate on the linguistic structure of the network, a limitation obviously comes from the language of the author. Speaking French, he is better able to explore this part of the network than another, something that could artificially produce a high clustering of his own linguistic community. This risk is minor here due to a special effort made to find a maximum of users representing the linguistic diversity of the field, particularly in German-, Italian- and Spanish-speaking areas.

The fact that the list is public is likely to skew the results of this analysis: conscious of having been added, some users could use it to discover and follow new users, which would have the effect of increasing the network's density. Similarly, we cannot exclude that the process has led some to discover the list author's account: they may have found the initiative or the profile interesting and will have therefore followed it, which could have caused a slight upgrade of the latter in the ranking. In the longer term, a public list is problematic because it is likely, gradually acquiring the status of "reference", to encourage compulsive subscription behaviours, such as users hoping to be "followed back" by colleagues. In itself, this behaviour is not a problem, it is a networking strategy that can be justified to socialise in a given community, but to use only one list for this is problematic: the more it is used for this purpose, the denser the network becomes, the more it impoverishes the diversity possibility in the field. But on the other hand, keeping this list public is mostly a way of giving the community a chance to discover unknown profiles and is a contribution to the friendly spirit of this social media. This also allows other researchers to use this corpus to conduct other types of studies: content analysis, interactions, biographies, shared links, etc.

Also note that the representativeness of Twitter is widely debated (Mislove, Jorgensen, Ahn, Onnela, & Rosenquist, 2011; Sloan et al., 2013), and that it is established that the social network's users are not a sample image of the population (Duggan, Ellison, Lampe, Lenhart, & Madden, 2015; Miller, Ginnis, Stobart, Krasodomski-Jones, & Clemence, 2015). While this representativeness is crucial to draw political conclusions (Boyadjian, 2014; Vainikka & Huhtamäki, 2015), the universities' landscape and the digital humanities are themselves such a little representation of the population that these considerations are difficult to apply here. Hence, the need to combine our analysis to a qualitative survey of these areas to assess this very special representativeness.

9. Conclusion

In this paper, we found that defining digital humanities as a "community" avoids endless debates on its disciplinary boundaries but does not allow us to know who's practicing them today. As an attempt to identify this field, leaving aside the epistemological discussion, our study shows that this item is analysable through a social media widely used by the so-called "digital humanists" (2,500 users). In analysing the network of "who's following who?", it was found that a small number of individuals and institutions are focusing so much attention that the graph appears to be very homogeneous

around them. The fact remains that many types of behaviour can be deduced from the graph and that structural characteristics of the network enable us to highlight some users holding remarkable positions. Specifically, we showed that French-speaking users, and to a lesser extent German-speaking users, stand out: the language factor strongly influences the network structure.

Obviously, any quantification leads to a form of objectification whose limits we need to understand. But we note that the availability of this type of data-set and the opportunities offered by tools and theories such as social network analysis allows us to shed new light on this community.

Acknowledgements
Martin Grandjean is thankful to Vincent Barbay (Ecole Polytechnique Fédérale de Lausanne) for his help in the data recovery and processing and to Etienne Guilloud (Université de Lausanne) for his help correcting his English. The idea to analyse this kind of data-sets was initiated with Yannick Rochat (Ecole Polytechnique Fédérale de Lausanne) during the Pegasus Data works in 2012. Grandjean is also thankful to Frédéric Clavert (Université de Lausanne) and Olivier Le Deuff (Université Bordeaux Montaigne) for exchanging about the French position.

Funding
The author received no direct funding for this research.

Author details
Martin Grandjean[1]
E-mail: martin.grandjean@unil.ch
ORCID ID: http://orcid.org/0000-0003-3184-3943
[1] Department of History, University of Lausanne, Lausanne, Switzerland.

Citation information
Cite this article as: A social network analysis of Twitter: Mapping the digital humanities community, Martin Grandjean, *Cogent Arts & Humanities* (2016), 3: 1171458.

Notes

1. "Big Tent Digital Humanities", title of the *Digital Humanities 2011*, ADHO Conference, Stanford 2011.
2. http://www.twitter.com.
3. Data visualisation produced with Gephi (Bastian, Heymann, & Jacomy, 2009).
4. Not forgetting that the "closed world" impression is accentuated by the fact that the data-set is itself finite. The term is rather to be taken metaphorically.
5. We develop this typology between *demonstration* and *research* visualisation in (Grandjean, 2015a).
6. Linguistic associations: French, German, Spanish, Portuguese and Japanese; regional associations: European, Nordic, Australasian, South America, Argentinian and Israeli.
7. Global Outlook:Digital Humanities http://www.globaloutlookdh.org.
8. This choice of vocabulary should not make us forget that this "virtual" world is contained within the "real" world, and this especially as the use of online social networking is increasingly becoming a factor of scientific socialisation, trade and promotion.
9. Force Atlas 2 (Jacomy, Venturini, Heymann, & Bastian, 2014).
10. http://www.digitalhumanitiesnow.org.
11. http://www.dhsi.org.

References

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*, 361–362.

Beguerisse-Diaz, M., Garduno-Hernandez, G., Vangelov, B., Yaliraki, S. N., & Barahona, M. (2014). Interest communities and flow roles in directed networks: The Twitter network of the UK riots. *Journal of the Royal Society Interface*, 0940. doi:10.1098/rsif.2014.0940

Boyadjian, J. (2014). Twitter, un nouveau "baromètre de l'opinion publique"? *Participations*, 55–74. http://dx.doi.org/10.3917/parti.008.0055

Casilli, A., & Tubaro, P. (2012). Social media censorship in times of political unrest—A social simulation experiment with the UK riots. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 115*, 5–20. http://dx.doi.org/10.1177/0759106312445697

Darmon, D., Omodei, E., & Garland, J. (2015). Followers are not enough: A multifaceted approach to community detection in online social networks. *PloS ONE. 10,* e0134860. doi: 10.1371/journal.pone.0134860

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., & Madden, M. (2015). *Social media update 2014*. Pew Research Center. Retrieved from http://www.pewinternet.org/2015/01/09/social-media-update-2014/

Freeman, L. (1978). Centrality in social networks: Conceptual clarification. *Social Networks, 1*, 215–239. http://dx.doi.org/10.1016/0378-8733(78)90021-7

Grandjean, M. (2015a). Introduction à la visualisation de données: l'analyse de réseau en histoire. *Geschichte und Informatik, 18*, 109–128.

Grandjean, M. (2015b). *THATCamp geography*. Retrieved from http://www.martingrandjean.ch/data-visualization-thatcamp-geography/

Grandjean, M., & Rochat, Y. (2014). *The digital humanities network on Twitter (#DH2014)*. Retrieved from http://www.martingrandjean.ch/dataviz-digital-humanities-twitter-dh2014/

Hale, S. A. (2014). Global connectivity and multilinguals in the Twitter network. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 833–842.

Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology, 65*, 656–669. http://dx.doi.org/10.1002/asi.23101

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS ONE, 9*, 1–12.

Jussila, J., Huhtamaki, J., Henttonen, K., Karkkainen, H., & Still, K. (2014). Visual network analysis of Twitter data for co-organizing conferences: Case CMAD 2013. *System Sciences (HICSS)*, 1474–1483.

Milgram, S. (1967). The small-world problem. *Psychology Today, 2*, 60–67.

Miller, C., Ginnis, S., Stobart, R., Krasodomski-Jones, A., & Clemence, M. (2015). *The road to representivity, a Demos and Ipsos MORI report on sociological research using Twitter*. London: Demos.

Mislove, A., Jorgensen, S. L., Ahn, Y.-Y., Onnela, J.-P., & Rosenquist, J. N. (2011). Understanding the demographics of Twitter users. *International AAAI Conference on Weblogs and Social Media*, 554–557.

Moreno, J. L. (1934). *Who shall survive? A new approach to the problem of human interrelations*. Washington, DC: Nervous and Mental Disease Publishing.

Myers, S. A., Sharma, A., Gupta, A., & Lin, J. (2014). Information network or social network? The structure of the Twitter follow graph. *Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion*, 493–498.

Newman, M. J. (2010). *Networks*. Oxford: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199206650.001.0001

Rochat, Y. (2014). *Character networks and centrality* (Thesis). Lausanne: University of Lausanne.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: Real-time event detection by social sensors. *World Wide Web*, 851–860.

Sloan, L., Morgan, J., Housley, W., Williams, M., Edwards, A., Burnap, P., & Rana, O. (2013). Knowing the Tweeters: Deriving sociologically relevant demographics from Twitter. *Sociological Research Online*, 1–11.

Stepanyan, K., Borau, K., & Ullrich, C. (2010). A social network analysis perspective on student interaction within the Twitter microblogging environment. *The 10th IEEE International Conference on Advanced Learning Technologies*, 70–72.

Stieglitz, S., & Dang-Xuan, L. (2012). Political communication and influence through microblogging, an empirical analysis of sentiment in Twitter messages and retweet behavior. *System Science (HICSS)*, 3500–3509.

Subbian, K., & Melville, P. (2011). Supervised rank aggregation for predicting influencers in Twitter. *Social Computing*, 661–665.

Suh, B., Hong, L., Piroll, P., & Chi, E. H. (2010). Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. *Social Computing*, 177–184.

Terras, M. (2010). Present, not voting: Digital humanities in the panopticaon. *Closing Plenary Speech at DH2010*. Retrieved from http://melissaterras.blogspot.ch/2010/07/dh2010-plenary-present-not-voting.html

Vainikka, E., & Huhtamäki, J. (2015). Tviittien politiikkaa – poliittisen viestinnän sisäpiirit Twittterissä. *Media & viestintä*, 165–183.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*, 440–442. http://dx.doi.org/10.1038/30918

cogent oa

© 2016 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:
Share — copy and redistribute the material in any medium or format
Adapt — remix, transform, and build upon the material for any purpose, even commercially.
The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:
Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.
You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

***Cogent Arts & Humanities* (ISSN: 2331-1983) is published by Cogent OA, part of Taylor & Francis Group.**

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com