

Imaginaries and STS literatures citation analysis

```
In [1]: %load_ext autoreload
        %autoreload 2
```

```
In [2]: import pandas as pd
import cit_anal as ca
import pickle
import re
import operator
from pylab import *
from IPython.display import HTML
```

Literatures: imaginaries (all humanities/social science fields) and the sts literature (from 3 major journals)

The analysis explores the relation between **two literatures** - the imaginaries literature and the sts literature:

- I. For 'imaginaries,' the query was 'imaginaries or imaginary' in any social science or humanities publication (all years);
- II. For STS, the query asked for all publications from Social Studies of Science, STHV and Science as Culture (all years).

The queries were run on Thomson Scientific Web of Science. Full records including cited references were downloaded for both literatures.

The literatures

As it happens, there are roughly the same number of publications in both literatures.

```
In [3]: imdf = ca.load_records(dir = 'data/')
print(str(imdf.shape[0]) + ' imaginary references to June 2013')
3690 imaginary references to June 2013
```

```
In [4]: stsd = ca.load_records(dir = 'data_sts_authors/')
print(str(stsd.shape[0]) + ' references published in the major sts journals to June 2013')
3714 references published in the major sts journals to June 2013
```

Whereas the STS literature comes from 3 journals, the imaginaries literature comes from many journals and fields. Many of these fields overlap to some degree. Literature predominates, quite massively.

In [9]:

```
## to get simplified view

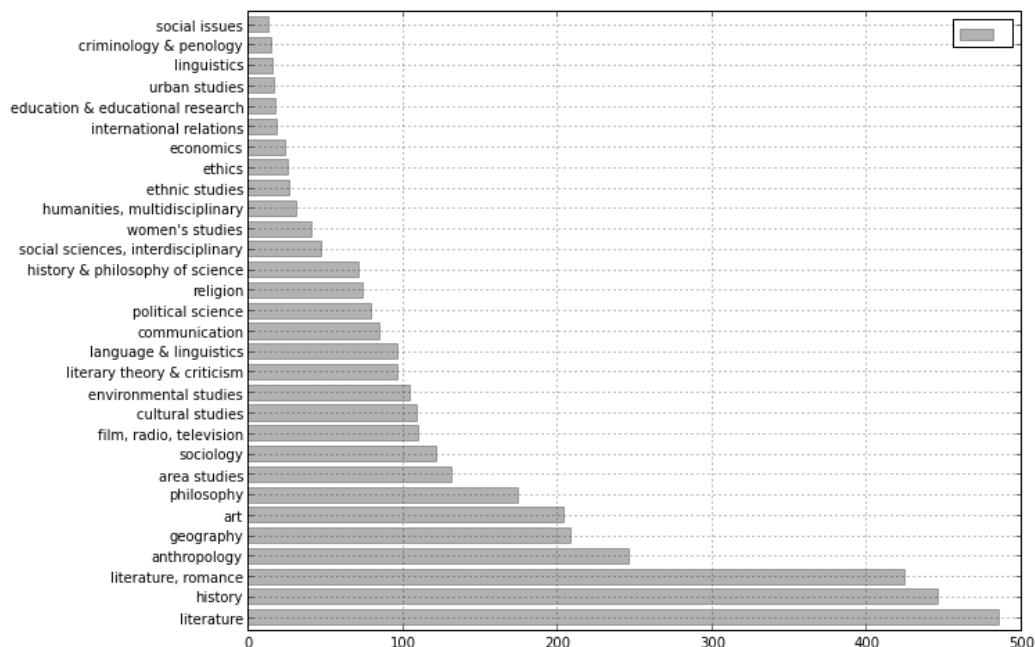
imdf = ca.field_normalise(imdf)

vc = imdf.field.value_counts()[0:30]
vc.plot(kind='barh', alpha=0.3, color='k', figsize=(10,8), legend=True)

print(imdf.WC.value_counts()[0:30])
```

Literature	474
Literature, Romance	425
History	367
Anthropology	172
Art	164
Philosophy	153
Geography	144
Sociology	112
Film, Radio, Television	97
Language & Linguistics; Literature	74
Religion	73
Environmental Studies; Geography	70
Cultural Studies	66
Political Science	62
Area Studies	57
Communication	49
Literary Theory & Criticism	44
Women's Studies	41
History & Philosophy Of Science	36
Social Sciences, Interdisciplinary	30
Literary Theory & Criticism; Literature	23
Art; Humanities, Multidisciplinary	23
Language & Linguistics; Literature, Romance	21
History; Religion	21
Area Studies; Humanities, Multidisciplinary	21
Environmental Studies; Urban Studies	18
Urban Studies	17
Humanities, Multidisciplinary; Social Sciences, Interdisciplinary	15
History; Literature, Romance	15
Geography; Planning & Development; Urban Studies	15

dtype: int64



Similarly the publication venues of for the imaginaries literature are quite diverse. In the list below, *Environment and Planning D* and *Theory, Culture & Society* stand out as places where STS articles might also appear. But basically, the imaginaries literature is in literature journals, and particularly the French ones (7 of the top 10).

```
In [10]: print(imdf.S0.value_counts()[0:50])
```

STUDI FRANCESI	66
FRENCH STUDIES	48
REVUE D HISTOIRE LITTERAIRE DE LA FRANCE	42
QUINZAINE LITTERAIRE	37
ENVIRONMENT AND PLANNING D-SOCIETY & SPACE	32
FRENCH REVIEW	31
MODERN LANGUAGE REVIEW	31
RLC-REVUE DE LITTERATURE COMPAREE	29
GEOFORUM	27
AMERICAN HISTORICAL REVIEW	25
CONNAISSANCE DES ARTS	24
THEORY CULTURE & SOCIETY	24
ENVIRONMENT AND PLANNING A	23
ROMANTISME	22
CULTURAL STUDIES	22
REVISTA IBEROAMERICANA	19
RIVISTA DI LETTERATURE MODERNE E COMPAREE	18
HOMME	18
WORLD LITERATURE TODAY	18
SOCIAL & CULTURAL GEOGRAPHY	17
CINEFORUM	17
HISTORIA MEXICANA	17
OEIL-MAGAZINE INTERNATIONAL D ART	16
ETUDES ANGLAISES	16
SODOM ALS SYMPTOM: GLEICHGESCHLECHTLICHE SEXUALITAT IM CHRISTLICHEN IMAGINAREN:	
RELIGIONSGESCHICHTLICHE ANAMNESE. A HISTORY OF RELIGION	16
NOUVELLE REVUE FRANCAISE	16
SOCIOLOGICAL REVIEW	15
NINETEENTH-CENTURY FRENCH STUDIES	15
INTERNATIONAL JOURNAL OF URBAN AND REGIONAL RESEARCH	15
BULLETIN OF HISPANIC STUDIES	15
KUNSTWERK	15
REVISTA DE CRITICA LITERARIA LATINOAMERICANA	15
TEXTUAL PRACTICE	15
AMERICAN ETHNOLOGIST	15
POSITIF	15
CULTURAL GEOGRAPHIES	14
ETUDES LITTERAIRES	13
AMERICAN ANTHROPOLOGIST	13
ANGLIA-ZEITSCHRIFT FUR ENGLISCHE PHILOLOGIE	13
REVUE HISTORIQUE	12
ANTIPODE	12
POETICS TODAY	12
PUBLIC CULTURE	12
ZEITSCHRIFT FUR ANGLISTIK UND AMERIKANISTIK	12
DIOGENES	12
HISTORIA	12
HISTORY OF THE HUMAN SCIENCES	11
SOCIAL & LEGAL STUDIES	11
AMERICAN BOOK REVIEW	11
SCIENCE-FICTION STUDIES	11

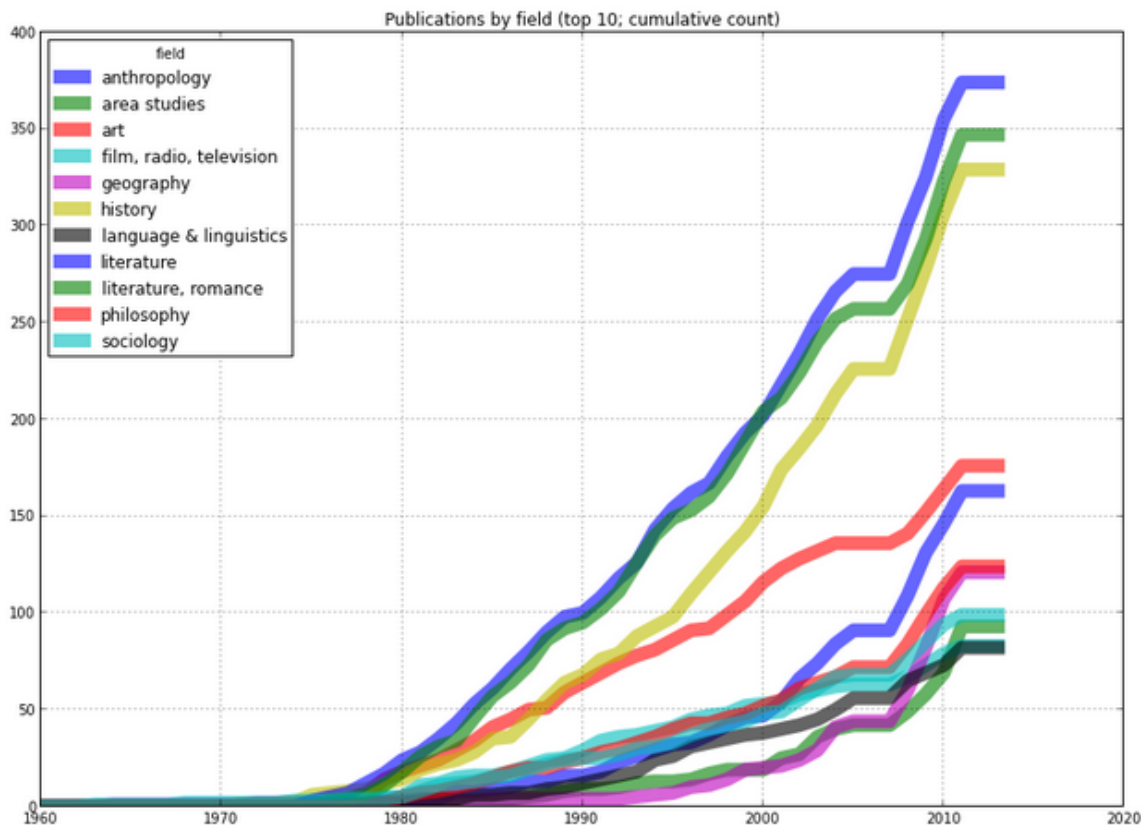
dtype: int64

Does the disciplinary mix of the imaginaries literature change over time?

In [11]: *#simplify the fields a bit -- they are too complicated*

```
fields_df = ca.fields_over_time(imdf)
cols = [col for col in fields_df.columns if fields_df[col].sum()>80]
fields_df[cols].cumsum().plot(alpha=0.6, legend=True, figsize=(14,10), lw=10, title='Publications
```

Out[11]: <matplotlib.axes.AxesSubplot at 0x75f8c90>



The imaginaries literature has grown most steeply in literature, literature (romance languages) and history, especially since 2000. These graphs are *not* normalised by discipline size. That is, they are perhaps many more publications in literature and history in general than say anthropology or sociology. Nearly all fields seem to show a much steeper growth since around 2005.

There is no point graphing this for the STS literature -- it is all code by a three fields, which is quite striking in itself. Why is it coded so narrowly by the Web of Science?

In [12]: `stsdw.WC.value_counts()`

```
Out[12]: Social Issues          1819
History & Philosophy Of Science 1774
Cultural Studies; History & Philosophy Of Science 121
dtype: int64
```

Who are the main authors in the two literatures, and who publishes in both literatures?

Given its multi-disciplinary span, I don't expect the imaginaries literature to have any major authors, but there might be some key figures worth looking at. It has **3900** unique authors.

The STS literature is much more coherent. It only has **2700** unique authors. Much more co-authorship appears here too.

```
In [13]: im_au = ca.author_list(imdf)
im_au_unique = set(im_au)
print('Number of authors in imaginaries literature: '+ str(len(im_au)))
print('Number of unique authors in imaginaries literature: '+ str(len(im_au_unique)))
im_au_counts = {au:im_au.count(au) for au in im_au_unique}
print(sorted(im_au_counts.iteritems(), key = operator.itemgetter(1), reverse=True)[0:20])
print('\n\n')
sts_au = ca.author_list(stsd)
sts_au_unique = set(sts_au)
print('Number of authors in sts literature: '+ str(len(sts_au)))
print('Number of unique authors in sts literature: '+ str(len(sts_au_unique)))
sts_au_counts = {au:sts_au.count(au) for au in sts_au_unique}
print(sorted(sts_au_counts.iteritems(), key = operator.itemgetter(1), reverse=True)[0:20])
```

```
Number of authors in imaginaries literature: 4170
Number of unique authors in imaginaries literature: 3868
[(['anonymous'], 41), ('brinkschroeder, michael', 15), ('nan', 9), ('comuzio, e', 7),
('sandiford, keith', 6), ('cesbron, g', 6), ('minca, claudio', 5), ('decaux, a', 4),
('prytherch, david l.', 4), ('pernot, d', 4), ('nahin, paul j.', 4), ('laurie, nina', 3),
('erni, john nguyet', 3), ('chen, fanfan', 3), ('kanceff, e', 3), ('jasanoff, sheila', 3),
('campbell, j', 3), ('lacroix, jean', 3), ('routledge, paul', 3), ('battaglia, d', 3)]

Number of authors in sts literature: 4825
Number of unique authors in sts literature: 2738
[(['anonymous'], 525), ('collins, hm', 33), ('lynch, m', 27), ('lafollette, mc', 18),
('mulkay, m', 18), ('pinch, t', 17), ('mackenzie, d', 17), ('gorman, me', 15), ('bloor, d',
15), ('woolgar, s', 15), ('yearley, s', 15), ('lynch, michael', 15), ('fuller, s', 14),
('jasanoff, s', 14), ('edge, d', 14), ('martin, b', 13), ('nelkin, d', 13), ('levidow, l',
12), ('barnes, b', 12), ('restivo, s', 12)]
```

The intersection between the author lists in both literatures is quite small. Only **52** authors publish in both:

```
In [14]: # overlap between author list

intersect = im_au_unique.intersection(sts_au_unique)
print(str(len(intersect)) + ' authors appear in both literatures')
print(intersect)

52 authors appear in both literatures
set(['sorensen, estrid', 'walker, mb', 'prince, ruth j.', 'prasad, amit', 'gottweis, h',
'smith, d', 'rapp, rayna', 'johnson, p', 'kellner, d', 'levidow, les', 'curtis, b',
'waterton, c', 'geissler, p. wenzel', 'bayma, t', 'manderscheid, katharina', 'greenberg, m',
'hyysalo, sampsa', '[anonymous]', 'star, sl', 'taylor, peter j.', 'woehrer, veronika', 'moon,
s', 'miller, c', 'geschiere, p', 'schmaus, w', 'smith, elta', 'cohen, l', 'chambers, r',
'harrasser, doris', 'wynne, b', 'kim, sang-hyun', 'cutting, andrew', 'rosen, p', 'oreskes,
naomi', 'jasanoff, sheila', 'mazlish, b', 'pels, d', 'durbin, pt', 'aitken, mhairi',
'townsend, christopher', 'campbell, elaine', 'bowker, gc', 'koch, gertraud', 'fraser,
suzanne', 'adams, c', 'reardon, jenny', 'papadopoulos, dimitris', 'tomaselli, s', 'sanders,
s', 'pickersgill, marty', 'fortun, m', 'porter, r'])
```

Cited references in the imaginaries literature

The aim here is to see:

- I. what references are cited?
- II. what references are cited most often?

The references cited give some idea of what the imaginaries literature is drawing on, and how it relates to different social science and humanities fields.

What references are cited in the imaginaries literature?

```
In [15]: cited_refs = imdf.cited_refs.tolist()

cited_refs_full = [r for ret in imdf.cited_refs for r in ret]

# the unique set of cited refs
cited_ref_set={r for ret in imdf.cited_refs for r in ret}
print('There are ' + str(len(cited_ref_set)) + ' unique references cited')

There are 88074 unique references cited
```

```
In [149]: # DON'T RUN THIS UNLESS NECESSARY - it takes ages
#to see what is cited most -- this takes a couple of minutes, so I'm saving it
ref_counts={ref:cited_refs_full.count(ref) for ref in cited_ref_set}
pickle.dump(ref_counts, file=open('imref_counts.pyd', 'w'))
```

```
In [16]: # to look at range of values, easier in a series
imref_counts = pickle.load( file=open('imref_counts.pyd', 'r'))
ref_counts_s=pd.Series(imref_counts.values(), index = imref_counts.keys())
ref_counts_s = ref_counts_s.dropna().order(ascending=False)
print('There are ' + str(sum(ref_counts_s >1)) + ' references cited more than once')

There are 4950 references cited more than once
```

Who is cited most in the imaginaries literature?

Looking at the citation counts, the most cited authors are pretty much expected - Anderson, Appadurai, Butler, Said, etc. They are 'big names' in general in the humanities and social science. Some of the most cited come from the 'imaginary' world: Anderson, Taylor, Castoriadis, Gaonkar, Lacan, etc. There are a few sts figures here too: Latour and Haraway. I will explore their presence below.

```
In [17]: print(ref_counts_s[ref_counts_s>10][1:50])
```

```
anderson benjamin, 1991, imagined communities      82
appadurai arjun, 1996, modernity large cult        57
taylor charles, 2004, modern social imagin         48
anderson b., 1983, imagined communities           47
castoriadis cornelius, 1987, imaginary i soc       43
bhabha h. k., 1994, location culture               39
lefebvre henri, 1991, production space            39
said edward w., 1978, orientalism                 38
hardt m., 2000, empire                             31
latour b., 1993, we have never been m            27
zizek s., 1989, sublime object ideol             25
haraway donna j., 1991, simians cyborgs wome      25
habermas jurgen, 1989, structural transform        25
bakhtin m., 1981, dialogic imagination            23
butler j., 1990, gender trouble femin             22
harvey david, 1989, condition postmodern          22
massey d, 1994, space place gender                22
said e. w., 1979, orientalism                     22
taylor c, 2002, public culture, v14, p91, doi 10.1215/08992363-14-1-91  22
gilroy paul, 1993, black atlantic moder           22
mcclintock anne, 1995, imperial leather rac       21
sibley david, 1995, geographies exclusio         21
bourdieu p, 1977, outline theory pract            21
geertz c., 1973, interpretation cultu            20
massey d., 2005, space                             20
giddens a., 1991, modernity self ident            19
clifford james, 1997, routes travel transl        19
ong a., 1999, flexible citizenship                18
laclau ernesto, 1985, hegemony socialist s        18
beck u, 1992, risk soc new moderni                18
gaonkar dp, 2002, public culture, v14, p1, doi 10.1215/08992363-14-1-1  17
butler j., 1993, bodies matter discurs            17
deleuze g, 1987, 1000 plateaus capita             17
said edward, 1993, culture imperialism            17
decerteau m, 1984, practice everyday li          15
pratt mary louise, 1992, imperial eyes travel     15
bourdieu pierre, 1984, distinction social c       15
harvey d, 1996, justice nature geogr              15
brah a., 1996, cartographies diaspo              15
williams raymond, 1973, country city              15
gregory derek, 1994, geographical imagina         14
harvey d, 2005, brief hist neoliberal            14
gramsci a, 1971, selections prison no             14
fabian johannes, 1983, time other anthr mak       14
lacan j, 1966, ecrits                             14
de certeau m, 1984, practice everyday li         14
latour b., 1999, pandoras hope essays            13
lacan jacques, 1977, ecrits                      13
escobar a, 1995, encountering dev mak            13
dtype: int64
```

For the STS literature

Now ask the same questions of the STS literature

- I. what references are cited?
- II. what references are cited most often?

What is cited in the STS literature?

```
In [18]: stscited_refs = stsd.cited_refs.tolist()

stscited_refs_full = [r for ret in stscited_refs for r in ret]

# the unique set of cited refs
stscited_ref_set = {r for ret in stscited_refs for r in ret}
print('There are ' + str(len(stscited_ref_set)) + ' unique references cited in the STS literature')

There are 77624 unique references cited in the STS literature
```

```
In [153]: ##DO NOT RUN UNLESS NECESSARY - takes a while. Go straight to next cell
#to see what is cited most -- this takes a couple of minutes, so I'm saving it - only run once
stsref_counts = {ref: stscited_refs_full.count(ref) for ref in stscited_ref_set}
pickle.dump(stsref_counts, file=open('stsref_counts.pyd', 'w'))
```

```
In [19]: # to look at range of values, easier in a series
stsref_counts = pickle.load( file=open('stsref_counts.pyd', 'r'))
stsref_counts_s = pd.Series(stsref_counts.values(), index = stsref_counts.keys())
stsref_counts_s = stsref_counts_s.dropna().order(ascending=False)
print('There are ' + str(sum(stsref_counts_s > 1)) + ' references cited more than once')

There are 11388 references cited more than once
```

Who is cited most in the STS literature?

As the table below shows, the most cited authors in STS are much more clearly STS. This is a pretty inward focused literature, with only a few outsiders showing up -- Beck, and perhaps Garfinkel. No Butlers or Appadurai's in the A-list citations. Actually, judging by this analysis, the field of STS should be called 'Latour studies.'

In other ways, the STS literature is broader. Although imaginaries cites 88000 references and STS 76000, which is roughly commensurate, STS cites **11388** or so references more than once whereas the imaginaries literature only cites **4300**. This is despite the fact that they are roughly the same size. This suggests that STS actually draws on a wider span of publications than the imaginaries literature.

More importantly, the citation counts are massively higher in the STS literature. (Collins 2002) is higher than Appadurai in the imaginaries literature. Despite the breadth of its cited literature, STS is a much more focused literature. It cites the same references more often.


```
In [20]: print(pd.DataFrame(stsref_counts_s[stsref_counts_s>10]).head(50))
```

nan	0
latour b., 1987, sci action follow sc	446
latour b., 1987, sci action	183
latour b., 1979, laboratory life soci	112
pinch t., 1987, social construction	96
star sl, 1989, soc stud sci, v19, p387, doi 10.1177/030631289019003001	91
collins hm, 1985, changing order repli	89
bloor d, 1976, knowledge social ima	86
gilbert g. n., 1984, opening pandoras box	81
callon m., 1986, power action belief, p196	72
gieryn tf, 1983, am sociol rev, v48, p781, doi 10.2307/2095325	68
karin knorr-cetina, 1981, manufacture knowledg	66
latour b., 1993, we have never been m	64
shapin s., 1985, leviathan air pump h	64
haraway donna j., 1991, simians cyborgs wome	62
epstein s., 1996, impure sci aids acti	62
kuhn ts, 1970, structure sci revolu	61
gieryn thomas f., 1999, cultural boundaries	61
trawweek s., 1988, beamtimes lifetimes	61
collins hm, 1975, sociology, v9, p205, doi 10.1177/003803857500900202	61
latour b., 1988, pasteurization franc	60
haraway d, 1997, modest witness 2 mil	60
collins hm, 2002, soc stud sci, v32, p235, doi 10.1177/0306312702032002003	57
latour bruno, 1986, lab life constructio	56
hughes t., 1983, networks power elect	54
pinch tj, 1984, soc stud sci, v14, p399, doi 10.1177/030631284014003004	54
bowker gc, 1999, sorting things out c	53
akrich m., 1992, shaping technology b, p205	53
gibbons m., 1994, new production knowl	52
porter t. m., 1995, trust numbers pursui	52
kuhn ts, 1962, structure sci revolu	52
mol a., 2002, body multiple ontolo	51
latour b., 1983, sci observed perspec, p141	49
latour b, 1979, lab life social cons	49
jasanoff s, 1990, 5 branch sci adviser	48
collins hm, 1974, sci stud, v4, p165, doi 10.1177/030631277400400203	48
beck u, 1992, risk soc new moderni	48
latour b., 1999, pandoras hope essays	47
jasanoff sheila, 2005, designs nature sci d	46
fujimura jh, 1987, soc stud sci, v17, p257, doi 10.1177/030631287017002003	46
garfinkel h., 1967, studies ethnomethodo	45
bijker w., 1995, bicycles bakelites b	44
callon m., 1986, mapping dynamics sci	43
shapin steve, 1985, leviathan air pump	43
collins hm, 1981, soc stud sci, v11, p3, doi 10.1177/030631278101100101	41
pickering a, 1995, mangle practice time	40
suchman lucy a., 1987, plans situated actio	39
mackenzie d, 1985, social shaping techn	39
mulkay mj, 1976, soc sci inform, v15, p637, doi 10.1177/053901847601500406	39
winner l, 1993, sci technol hum val, v18, p362, doi 10.1177/016224399301800306	39

The intersection between the STS and the imaginaries literature

A couple of questions here:

- I. What is intersection between the most-cited references in STS and imaginaries?
- II. How does STS literature cite the imaginaries literature?
- III. How is the STS literature present in the imaginaries literature?

Who do they both cite? The intersection between STS and imaginaries in terms of who they both cite ('co-citation')

The overlap in the cited references is around **2500** references.

```
In [21]: print(str(len(cited_ref_set.intersection(stscited_ref_set))) + ' references are found in the citer
2456 references are found in the cited references in both literatures
```

Here is the list of references that are common to both -- too long to look at properly. It'd be good to sort them by order of importance -- e.g. how

often they are cited.

```
In [22]: stsim_cocit = sorted(list(cited_ref_set.intersection(stscited_ref_set)))
stsim_cocit[100:120]
```

Given this is too long to look at, we might be able to get a rough idea by counting the surnames -- this is rough, but gives an idea of where the overlap lies:

```
In [23]: author_refs = [ref.split(' ')[0] for ref in stsim_cocit]
author_refs_set = set(author_refs)
au_counts = {au:author_refs.count(au) for au in author_refs_set}
stsim_cited_intersection = sorted(au_counts.iteritems(), key=operator.itemgetter(1), reverse=True)

print('There are ' + str(len(stsim_cited_intersection)) + ' references co-cited between the imag:
print(pd.DataFrame(stsim_cited_intersection[0:100], columns=['author', 'count']).head(30))
```

There are 1397 references co-cited between the imaginaries and STS literature

	author	count
0	foucault	42
1	latour	37
2	bourdieu	15
3	beck	14
4	law	14
5	jasanoff	14
6	mackenzie	13
7	habermas	13
8	strathern	13
9	williams	12
10	haraway	12
11	brown	11
12	derrida	11
13	thompson	11
14	rose	11
15	de	10
16	goffman	10
17	callon	10
18	butler	10
19	smith	10
20	rabinow	10
21	millar	9
22	wynne	9
23	thrift	9
24	martin	9
25	van	9
26	cooper	8
27	anderson	8
28	taylor	8
29	clarke	8

There are **1497** references co-cited by the imaginaries and STS literature. Again, this is somewhat rough because it is just using surnames. These references could be grouped in different ways, but a couple of features stand out.

- The philosophy intersection -- Foucault, Derrida, Marx, Deleuze, Wittgenstein, Heidegger, Baumann, Benjamin, Zizek, Butler
- The sociology intersection -- Bourdieu, Goffman, Beck, Williams, Castells, Merton, Weber, Baudrillard, Luhmann, Boltanski
- The political theory intersection -- Habermas, Laclau, Mouffe, Taylor, Dean
- The STSish intersections -- Latour, Law, Callon, Jasanoff, Wynne, Rheinberger, Mackenzie, Suchman, Pickering, Woolgar, Barry, Bijker, Collins
- The feminist intersection -- Butler, Haraway
- The anthropology intersection -- Strathern, Rabinow, Franklin, Appadurai, Comaroff, Geertz, Lock
- The historian/philosopher of science intersection -- Kuhn, Hacking, Porter, Young, Daston
- The media/cultural studies intersection -- Hall, Williams, Haraway, Keller

How is the STS literature included in the imaginaries literature?

This question can be answered a couple of different ways:

- the STS literature is a part of the imaginaries literature, but only a small part judged by journal articles. Of the 3500 imaginaries publications, only a dozen or so appear in STS journals.

```
In [24]: # query for journal SOC STUD SCI, SCI TECHNOL HUM VAL, SCIENCE CULTURE in the imaginaries dataset
sss = 'SOC STUD SCI|Social Studies of Science'
sthv = 'SCI TECHNOL HUM VAL|SCIENCE TECHNOLOGY & HUMAN VALUES'
sac = 'SCIENCE AS CULTURE'
patn = re.compile('|'.join([sss, sthv, sac]), re.IGNORECASE)

journals = imdf.S0.tolist()

l=[i.group() for i in [patn.search(str(s)) for s in journals] if i is not None]
print(str(len(l)) + ' STS journal articles appear in the imaginaries literature')
```

10 STS journal articles appear in the imaginaries literature

- How does the imaginaries literature cite authors who 'belong' to sts? There are around 2500 STS authors in the STS literature. How are they cited amidst the 88000 references cited in the imaginaries literature?

```
In [25]: # set of unique sts authors
stsau = list(ca.author_list(stsdf, True))

#take out commas
stsau_pattern= ['('+au.replace(',','')+')' for au in stsau]

# regular expressions can only deal with 100 at a time, so break pattern into chunks
au_patterns = [re.compile('|'.join(stsau_pattern[i:i+99])) for i in xrange(0, len(stsau_pattern),
```

```
In [16]: # search for all sts literature authors in the cited references of the imaginaries literature
# this search takes a while as it is searching for how often 2500 authors appear in the 88000 cit
res_full=[au_pattern.search(ref) for ref in cited_refs_full for au_pattern in au_patterns]
```

```
In [56]: # count how often they appear in the cited refs for imaginaries literature
sts_authors_cited = sorted([r.group() for r in res_full if r is not None and len(r.group())>1])
# load pickle below to save time
```

...

```
In [26]: # has been pickled to save time
# if rerunning, need to pickle again, etc.
#pickle.dump(sts_authors_cited, file=open(name='sts_cited_authors_cited.pyd', mode='wb'))
sts_authors_cited = pickle.load(file=open('sts_cited_authors_cited.pyd', mode='rb'))
```

```
In [27]: print('There are ' + str(len(sts_authors_cited)) + ' references to STS authors in the imaginaries
There are 2564 references to STS authors in the imaginaries literature
```

The author names need to be cleaned up for duplicates, etc, but it is fairly clear that the imaginaries literature cites the STS literature along similar lines fairly heavily. Some of the authors in the STS literature are not particularly STSish (e.g. Sherry Turkle or Mary Douglas or Douglas Kellner).

More importantly, it is fairly clear that only a **relatively small proportion** of the STS literature enters into the imaginaries literature. The total STS references amount to **2500** of the 88000, and the number of authors who actually get cited more than a few times account for the vast majority of citations of STS literature in the imaginaries literature (1800 out of 2500) is quite small -- around **150** authors.

```
In [28]: sts_authors_cited_counts = {au:sts_authors_cited.count(au) for au in set(sts_authors_cited)}
sorted_sts_cited_authors = sorted(sts_authors_cited_counts.iteritems(), key=operator.itemgetter(1))

print(str(len(sorted_sts_cited_authors)) + ' STS authors are cited in the imaginaries literature')
print(str(len([a for a in sorted_sts_cited_authors if a[1]>=5])) + ' of these are cited 5 or more')
print(str(sum([cited[1] for cited in sorted_sts_cited_authors if cited[1] >=5])) + ' of the 2500')
print('\n\n')

print(pd.DataFrame(sorted_sts_cited_authors[0:30], columns=['author', 'count']).head(30))
```

```
526 STS authors are cited in the imaginaries literature
145 of these are cited 5 or more times
1846 of the 2500 references from STS come this latter group
```

	author	count
0	latour b	188
1	law j	79
2	berg m	53
3	smith d	50
4	taylor p	44
5	douglas m	39
6	ney s	33
7	callon m	33
8	jasanoff s	30
9	mann g	27
10	smith r	25
11	martin e	24
12	brown p	23
13	mackenzie d	23
14	wynne b	21
15	geschiere p	21
16	bal r	20
17	taylor pj	19
18	nandy a	19
19	lee s	18
20	smith e	18
21	kellner d	18
22	latour bruno	17
23	cohen l	17
24	johnston r	17
25	hard m	16
26	martin p	16
27	lee j	15
28	turkle s	15
29	williams b	15

These authors are not the quite the same authors who appeared in the intersection between the imaginaries and STS citations. Here we are looking at how often people publishing in STS journals are cited in imaginaries literature. Some of the names are the same -- Latour, Law, Mackenzie, Wynne -- but many of the top names in the intersection between the cited references do not appear here -- Haraway, Foucault, Strathern -- mainly because they have never published in STS journals.

There are also names here -- Marc Berg, Emily Martin -- that didn't appear in the imaginaries-STC intersection at all. These are cases where STS work is being taken into the wider imaginaries literature.

How does the STS literature cite the imaginaries literature?

This is a key question. To get a rough idea, look for which of the 3868 imaginaries authors are cited in the 76000 references cited in sts literature.

```
In [29]: # set of unique sts authors
imau = list(ca.author_list(imdf, True))

#take out commas
imau_pattern= ['('+au.replace(',','')+')' for au in imau]

# regular expressions can only deal with 100 at a time, so break pattern into chunks

imau_patterns = [re.compile('|'.join(imau_pattern[i:i+99])) for i in xrange(0, len(imau_pattern),
print('There are ' + str(len(imau)) + ' unique authors in imaginaries literature')
```

```
There are 3868 unique authors in imaginaries literature
```

```
In [*]: imres_full=[imau_pattern.search(ref) for ref in stscited_refs_full for imau_pattern in imau_patte
```

Here is the list of imaginary authors cited in the sts literature. Some of them are STS people, since STS authors have contributed to the imaginaries literature, especially when they published outside the major STS journals.

```
In [*]: print(set(sorted([r.group() for r in imres_full if r is not None and len(r.group())>1])))
```

Breaking this down a bit further, if the STS literature cites around 7500 unique references, roughly 350 of these come from the imaginaries literature authors. Amongst this group, roughly 100 imaginaries authors account for most of the references from the STS literature. Judging from the list below, some of the key references are authors who overlap between STS and other fields.

```
In [140]: im_authors_cited = sorted([r.group() for r in imres_full if r is not None and len(r.group())>1])
print('There are ' + str(len(im_authors_cited)) + ' references to imaginary authors in the STS li
im_authors_cited_counts = {au:im_authors_cited.count(au) for au in set(im_authors_cited)}
sorted_im_cited_authors = sorted(im_authors_cited_counts.iteritems(), key=operator.itemgetter(1),
im_au_total = len(im_authors_cited)
print(str(len(sorted_im_cited_authors)) + ' individual imaginary authors are cited in the STS lite
print(str(len([a for a in sorted_im_cited_authors if a[1]>=5])) + ' of these are cited 5 or more t
print(str(sum([cited[1] for cited in sorted_im_cited_authors if cited[1] >=5])) + ' of the ' + str
print('\n')
print(sorted_im_cited_authors[0:30])
```

```
There are 4035 references to imaginary authors in the STS literature
364 individual imaginary authors are cited in the STS literature
105 of these are cited 5 or more times
3587 of the 4035 references from STS come this latter group
```

```
[('nan', 1054), ('wynne b', 410), ('star sl', 214), ('an g', 208), ('jasanoff sheila', 113), ('bo
66), ('smith d', 55), ('miller c', 48), ('berger h', 45), ('king a', 39), ('brown j', 38), ('hall
('white l', 26), ('gottweis h', 25), ('taylor c', 25), ('wright l', 24), ('porter r', 20), ('frie
19), ('mack j', 19), ('porter a', 18), ('waterton c', 18), ('davis d', 17), ('berman m', 17), ('h
15), ('joly p', 14), ('allen j', 14), ('thomas s', 14), ('deleuze gilles', 14), ('dean m', 13), ('
m', 12), ('adams j', 11), ('braun b', 11), ('cohen e', 11), ('thomas j', 11), ('rossi p', 11)]
```

What references are co-cited?

So far, we have been looking at how the imaginaries literature is present in the STS literature, and vice versa. We have looked at this by looking for STS authors in the cited references, and in the

Co-cited means that the same reference can be found cited in two 'different' publications. The fact that two publications both cite the same reference suggests that have some relation. Here I looked at the cited references for all the imaginaries publications, and ask for all the intersections between cited references. The bigger the intersection between the reference lists, the more substantial is the relationship. Perhaps.

```
In [23]: im_co = ca.co_citation(imdf)
sts_co = ca.co_citation(stsdf)
```

```
In [24]: print(str(len(im_co))+' co-citation sets exist in the imaginaries literature')
print(str(len(sts_co))+' co-citation sets exist in the STS literature')
```

```
58206 co-citation sets exist in the imaginaries literature
312270 co-citation sets exist in the STS literature
```

Again, the high co-citation count in STS literature shows that it is much more inwardly related than the imaginaries literature. The STS literature has roughly **five** times the number of co-citation sets as the imaginaries literature.

```
In [26]: # only run this if repeating the analysis somehow
pickle.dump(im_co, open('im_co_cit.pyd', 'w'))
pickle.dump(sts_co, open('sts_co_cit.pyd', 'w'))
```

```
In [20]: # to see size of co-citation sets in sts
sts_co = pickle.load(open('sts_co_cit.pyd', 'r'))
im_co = pickle.load(open('im_co_cit.pyd', 'r'))
#sorted([sts_co.count(set) for set in sts_co[0:2000]], reverse=True)[0:30]
```

```
In [28]: sts_co_set_lengths = sorted([len(s) for s in sts_co], reverse=True)
im_co_set_lengths = sorted([len(s) for s in im_co], reverse=True)
```

```
In [40]: print('Size of co-cited reference sets in STS literature:', sts_co_set_lengths[0:100:2])
print('Size of co-cited reference sets in imaginaries literature:', im_co_set_lengths[0:100:2])

('Size of co-cited reference sets in STS literature:', [223, 170, 147, 142, 115, 103, 99, 91,
88, 82, 79, 77, 73, 69, 68, 67, 66, 64, 62, 60, 58, 58, 57, 56, 54, 54, 54, 50, 49, 44, 42,
40, 40, 38, 36, 36, 34, 32, 32, 32, 31, 31, 30, 29, 29, 29, 29, 28, 28, 28])
('Size of co-cited reference sets in imaginaries literature:', [59, 27, 27, 25, 23, 23, 22,
18, 18, 17, 16, 14, 13, 13, 13, 12, 12, 11, 11, 11, 10, 10, 9, 9, 9, 9, 9, 9, 8, 8, 8, 8, 8,
8, 8, 8, 8, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7])
```

Not sure what to make of this, but again, the much greater number of co-citations, and the higher co-citation counts in the STS literature suggest that it is much more internally related than the imaginaries literature.