Machine Learning Methodologies and Large Data Text Corpora

Luke Barnesmoore,
University of British Columbia
luke.barnesmoore@geog.ubc.ca

Jeffery Huang, Laurent El Ghaoui
University of California at Berkeley
elghaoui@eecs.berkeley.edu

Abstract: With the rise of social media as a focal point for interaction in both global and local social communities, big data has become a key feature of social science research in the 21st century. As the size of corpora on sites like Facebook and Twitter have grown, a need has risen for more and more sophisticated computer science tools to collect data and identify statistical trends therein. After describing our integration of Netvizz scraping software with our Statnews.org language analysis software we provide a case study of our tool's application through analyzing Libertarian Facebook data within the lens of Barnesmoore's History of Assemblage Model (HoAM).

Introduction

One of the great challenges of modern social science is the task of acquiring and sifting through the deluge of information produced, circulated and consumed in the digital sphere each day. "According to computer giant IBM, 2.5 exabytes - that's 2.5 billion gigabytes (GB) - of data was generated every day in 2012."[1] Acquiring data is a costly and, or time consuming process: many of the large newspapers charge tens of thousands of dollars a year for access to their raw data; Twitter has recently acquired a firm 'Gnip' who will reportedly be charging $360,000 a year for access to 50% of Twitter's historical data;[2] scraping pages for data requires a team of skilled computer scientists and infrastructure to run the scraping program and store the data. Sifting through the data is difficult for obvious reasons. This paper, for example, analyzes a dataset derived from hundreds of thousands of comments by Libertarian Facebook pages and from tens of millions of comments on these page comments by Facebook users; obviously, such a data set would take years and years to sift through with any traditional means.

We (the UC Berkeley Statnews.org Lab)[3] have developed a tool that aims to circumvent these two essential problems. Our tool solves the first problem (acquisition of data) through integration with an existing app (Netvizz) that allows users to scrape data (comments by the page and user comments on the page

---

[1] Wall, M. (2014) "Big Data: Are You Ready for Blast-off?" *BBC News*, 3 Mar. 2014. http://www.bbc.com/news/business-26383058
[2] Kirkpatrick, M. (2010) "Titter to Sell 50% of All Tweets for $360k/Year Through Gnip" Readwrite.com, 17 Nov. 2010. http://readwrite.com/2010/11/17/twitter_to_sell_50_of_all_tweets_for_360kyear_thro (https://gnip.com/sources/twitter/historical/)
[3] See the following articles for a sampling of recent statnews.org projects:
Gawalt, B., Jia, J., Miratrix, L., El Ghaoui, L., Yu, B. and Clavier, S. (2010) Discovering word associations in news media via feature selection and sparse classification. In *Proc. 11th SACM SIGMM International Conference on Multimedia Information Retrieval*.
Miratrix, L., Jinzhu, J., Yu, B., Gawalt, B., El Ghaoui, L., Barnesmoore, L., Clavier, S., (2014) "Concise Comparative Summaries (CCS) of Large Text Corpora with a Human Experiment." The Annals of Applied Statistics 8 (1), 499--529. doi:10.1214/13-AOAS698. http://projecteuclid.org/euclid.aoas/1396966296.
Clavier, S., Barnesmoore, L., El Ghaoui, L., and Li, G. C. (2012) "All the News That's Fit to Compare," International Studies Association 53rd Annual National Convention San Diego: ISA 1-37. http://files.isanet.org/ConferenceArchive/bb6556e4811e410c9c524cc152ae68a5.pdf
Barnesmoore, L., El Ghaoui, L. (2013) "English vs. Arabic: Differing Al Jazeera Frames," International Studies Association 54th Annual National Convention San Francisco: ISA, 1-19. http://files.isanet.org/ConferenceArchive/e1419ee1b6af407090d3eab494f84379.pdf
Barnesmoore, L., El Ghaoui, L. Pham, V. (2014) "Media Frames & Histories of Assemblage," International Studies Association 55th Annual National Convention Toronto: ISA, 1-57. http://web.isanet.org/Web/Conferences/Toronto%202014/Archive/1d0fb7f3-ee28-443a-9535-e2fd1f95e1ff.pdf

comments) from Facebook pages without access to capital or computer science skills. Our tool solves the second problem (analyzing large data text corpora whose scope is beyond the pale of traditional analysis) through allowing the Netvizz output files to be loaded into the statnews.org server for machine learning analysis. With this new function, Statnews.org users will be able to both collect and analyze their data, free of charge, and without any computer science skills.

To illustrate the applicability of this new tool, we conduct an analysis of the 'most liked' Libertarian Facebook Pages: Americans for Limited Government,[4] Campaign for Liberty,[5] The CATO Institute,[6] The Libertarian Party,[7] Libertarianism.org,[8] The Ron Paul Institute for Peace and Prosperity[9] and The Libertarian Republic.[10] To conduct this analysis we apply Barnesmoore's History of Assemblage Model (HoAM),[11] a mixed methodological analytics developed to interrogate the relationship between socially normative ontological regimes (an assemblage of cosmological, teleological, ontological, epistemological, ethical-moral and aesthetic assumptions) and norms of thought, behavior and being in publics socialized by said ontological regime, to the visualizations of statistically dominant trends produced by the statnews.org software. In summary, we ask ourselves: 'what ontological regime rests implicitly in the terms isolated by the software, and how this regime is likely to manifest given the environmental context of Libertarians in the US?' Given the ethos of the HoAM and the rhetoric of the Libertarian movement, we focus this analysis on assumptions concerning the nature of freedom perpetuated by the statistically dominant narratives in the Libertarian social media sphere.

History of Assemblage Model

Barnesmoore's History of Assemblage Model (HoAM) is an analytic model designed to interrogate the relationship of ontological regimes and norms of thought behavior and being as well as the potentials for power relations produced therein (Barnesmoore 2014, Barnesmoore 2015).[12] Ontological regimes are the assemblage of cosmological, teleological, ontological, epistemological, ethical-moral and aesthetic assumptions that articulate the potential for one's conception of reality, production of knowledge, etc. In applying the HoAM to statnews.org visualizations, we interrogate the isolated terms within their embedded context in order to determine the ontological regime implicit in the analyzed discursive materials (large data text corpora) through use of a thought experiment. From here we observe the relationship between the ontological regime implicit in the terms isolated by our software and the norms of thought, behavior and being in publics socialized by the analyzed discursive materials as well as the potentials for power relations established therein through examining the relationship between the ontological regime implicit in the isolated terms and the existing cognitive environment (existing ontological regime(s), norms of thought behavior and being and power dynamics) of the socialized public. Taking the below case study as our example, we first outline the ontological regime implicit within the isolated terms in our query of the term 'freedom' in our Libertarian Facebook Data Set. From here we outline the existing cognitive environment of the libertarian public and analyze the likely implications of socialization within the ontological regime implicit in our isolated term for norms of thought, behavior and being as well as potential power dynamics therein.

The essentiality of the 'thought experiment' as our methodological foundation for analysis of the relationship between ontological regimes, language and socialized publics is elucidated by reading Deleuze's linguistic theory through the lens of traditional (pre-positivist) rationalism.

[4] Americans for Limited Government. Facebook. https://www.facebook.com/AmericansForLimitedGovernment
[5] Campaign for Liberty. Facebook Page. https://www.facebook.com/CampaignforLiberty
[6] CATO Institute. Facebook Page. https://www.facebook.com/CatoInstitute
[7] Libertarian Party. Facebook Page. https://www.facebook.com/libertarians
[8] Libertarianism.org. Facebook Page. https://www.facebook.com/Libertarianism.org
[9] Ron Paul Institute for Peace and Prosperity. Facebook Page. https://www.facebook.com/RonPaulInstitute
[10] The Libertarian Republic. Facebook Page. https://www.facebook.com/libertarianrepublic
[11] See the following articles for a sampling of recent applications of the developing History of Assemblage Model (HoAM):
Barnesmoore, L., El Ghaoui, L. Pham, V. (2014) "Media Frames & Histories of Assemblage," International Studies Association 55th Annual National Convention Toronto: ISA, 1-57. http://web.isanet.org/Web/Conferences/Toronto%202014/Archive/1d0fb7f3-ee28-443a-9535-e2fd1f95e1ff.pdf
Barnesmoore, L., Donoso, J., Claiver, S., El Ghaoui, L. (In Press) Machine Learning Methodologies: Histories of Asembalge and Representations of Women in the Bible. International Journal of Ciritical Cultural Studies.
[12] Barnesmoore, L., El Ghaoui, L. Pham, V. (2014) "Media Frames & Histories of Assemblage," International Studies Association 55th Annual National Convention Toronto: ISA, 1-57. http://web.isanet.org/Web/Conferences/Toronto%202014/Archive/1d0fb7f3-ee28-443a-9535-e2fd1f95e1ff.pdf
Barnesmoore, L., Donoso, J., Claiver, S., El Ghaoui, L. (In Press) Machine Learning Methodologies: Histories of Asembalge and Representations of Women in the Bible. International Journal of Ciritical Cultural Studies.

"If the objection is leveled that these specific features pertain to politics and not linguistics, it must be observed how thoroughly politics works language from within, causing not only the vocabulary but also the structure and all of the phrasal elements to vary as the order-words change. A type of statement can be evaluated only as a function of its pragmatic implications, in other words, in relation to the implicit presuppositions, immanent acts, or incorporeal transformations it expresses and which introduce new configurations of bodies. True intuition is not a judgment of grammaticality but an evaluation of internal variables of enunciation in relation to the aggregate of the circumstances."[13]

Intuition can be defined as the ability to integrate systems of dimensionally incommensurable variables into an assembled whole unattainable through use of the peripatetic mind (dimensional incommensurability can be understood in terms of the inability to inscribe the whole of the dimensional quality of a three dimensional object on a two dimensional plane and is most clearly articulated in Ouspensky's *Tertium Organum*); the cultivation of rationality consists of developing knowledge of the forms (aeons) that structure manifestation, and "intuition… takes what is known by Reason and grasps it in a single act of the mind".[14][15] Intuition, then, can be understood as an interference pattern between the unified forms upon which reason is founded and the difference, change, chaos, multiplicity, etc. of that which is interpreted through the use of intuition (of the created). Intuition, then, allows us to harmonize the change, difference, chaos, etc. of passing time and physical space (which rises from the dimensional quality of the prima materia) with the dimensionally incommensurable unity of the aeons that structure manifestation without stripping either of their essential dimensional quality because intuition itself is a manifestation of the interaction between these incommensurable planes of dimensional consistency.

Taking ontological regimes as aeons, individuals and groups (along with their existing cognitive environment) as the prima materia and norms of thought, behavior and being of socialized publics (as well as the linguistic symbols in which the ontological regimes are implicit) as the interference pattern that forms the plane of change, difference, chaos, etc., it is essential that we use a thought experiment to analyze their relationship via intuition as this is the only means by which we can study their interaction without stripping either of their essential dimensional quality. Thus, as language derives its meaning from the entire assemblage within which it is embedded, our mixed methodology is 'progressive' in its capacity to incorporate a larger section of the assemblage (both the linguistic assemblage, the assemblage of ontological regimes embedded in the linguistic assemblage, the cognitive environment of the socialized public, etc.) in our analysis of the relationship between ontological regimes embedded in large data text corpora and norms of thought, behavior and being in publics socialized by the analyzed corpora.

[13] Deleuze, G. (1987) *Thousand Plateaus*, University of Minnesota Press p. 83
[14] Descartes, *Discourse on the Method of Rightly Conducting one's Reason and Seeking Truth in the Sciences*, (2010 http://www.earlymoderntexts.com/pdfs/descartes1637.pdf)
Descartes, *Meditations on First Philosophy,* (2002) http://www.vahidnab.com/med.pdf
Descartes, *The Passions of the Soul*, http://www.earlymoderntexts.com/pdfs/descartes1649.pdf
[15] Nadler, Steven, "Baruch Spinoza", The Stanford Encyclopedia of Philosophy (Fall 2013 Edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/fall2013/entries/spinoza/

Unity: Uncreated   Interference Pattern: Unity and Difference in Unity   Multiplicity: Created

Aeon: Forms that Structure Manifestation → Manifest Reality ← Prima Materia: Matter Unconditioned by Form

Spirit: Consciousness → Soul: Mind ← Body

Ontological Regime → Norms of Thought, Behavior and Perception of Being ← Individual/ Group: Existing Cognitive Environment

∞   ∞

As explanation of the necessity of applying intuitive thought experiments inherently suffers from what Spinoza calls the poverty of language (the incommensurability of the dimensional quality of an idea with the dimensional quality of language making perfect inscription impossible in the same manner that it is impossible to perfectly inscribe a three dimensional figure onto a two dimensional plane), we attempt to elucidate the point with the above graphic. On the far left of the graphic we find the column of unity. In this column the objects exist in a state of static unity (prior to manifestation) and are thus incommensurable with the reality we experience through our sensory faculties in manifestation as humans. On the far right of the graphic we find the column of multiplicity. In this column the objects can be understood as vessel for the manifestation (creation) of the objects in the column of uncreated unity. The central column represents the interference pattern formed by the interaction of the left and right columns. The vertical lines with the infinity sign at their base represent the infinity membranes that demark the boundaries between incommensurable dimensional qualities. One point we must remember is that infinity is relative; for example, there may not actually be an infinite number of facts in the abstract sense, but there are an infinite amount of facts when viewed from the perspective of human life (in its limited cognitive capacities and its short lifespan) wherein it would be impossible to learn all the facts of the universe. As infinity is relative, and infinity membranes demark the boundaries between dimensionally incommensurable planes of dimensional consistency, dimensional incommensurability is itself relative. As such, while the unity of the aeons and the multiplicity of the material world (in its being structured by the dimensional quality of the prima materia) are incommensurable with each other, both are commensurable with the interference pattern formed between them. As such, to study objects that are dimensionally incommensurable with each other we must find the dimensional quality that lies in between the two planes of dimensional consistency and move from there. For knowledge formation, the interference pattern that bridges the unity of the aeons and the multiplicity of the world we experience is thought. As a result, if we wish to study the relationship of objects that exist in both the state of abstract unity (ontological regimes) and multiplicity (individuals and groups as well as their manifest cognitive environment) we must use thought experiments.

In such generalization, as a function of the change, motion, multiplicity, chaos, etc. that typifies the dimensional consistency of the plane of dimensional consistency we experience through our sensory faculties, we must fail to encompass the contingency and complex difference that actually describes the ontological regime, norms of thought behavior and being and produced power dynamics within the libertarian public. As such, the HoAM is, from a certain perspective, a *methodology of failure.* In the necessary failure of our attempts

to generalize that which is inherently contingent and complexly different, the *methodology of failure* illustrates the inherent tension of the interaction between unity (the first cause and its emanated aeonian forms which structure manifestation) and difference (passing time and physical space, the created) that forms the essence of reality. This illustration aims to produce an apocalyptic revelation of the fact that all we can know (about reality in the dimension of change, motion, multiplicity, chaos, etc.) is that we cannot know (he who knows everything knows nothing, but he who knows nothing knows everything).[16] It is from such doubt of our knowledge of manifest phenomena that we can, following Descartes, begin to develop rational knowledge of the forms (aeons) that structure manifestation and, through operationalization of this rational knowledge in a single movement of the mind (intuition), begin to form knowledge of the truth of the interaction between aeonian forms (here ontological regimes) and manifestations in passing time and physical space that are structured by the aeons. The failure of the method, then, attempts to cultivate the experience intended by Descartes in *The Meditations*.

Statnews.org Software

This section provides an overview of the mechanics by which the Statnews.org system creates a summary of large data text corpora. We take a large text corpus, break the corpus down into time slices (months, quarters, years), and extract keywords that co-occur most frequently with a query term for each time slice. We then present a sorted list of documents containing both the query term and the extracted keyword, with emphasis placed on documents that contain the extracted keyword more often. Visualizing these keywords side-by-side across time slices tells us what words were *statistically* descriptive of our corpus at a certain point in time. It also provides us with the documents that provide context for why the isolated terms are important for our understanding of this corpus.

There are a few important optimizations in our system that help improve results. The first method is stop-word removal - words that present no significant meaning such as "the", "that", and "this" are ignored in our corpus even though they co-occur the most with any query term. The second method hinges on normalizing data using the tf-idf statistic, which helps us correct for terms that naturally occur more often in our dataset. For example, the term "money" might occur abundantly throughout a dataset provided by the Financial Times, but this may not be a meaningful keyword. The tf-idf statistic increases as a term occurs more often in a document, but is penalized if that term occurs very often throughout the entire corpus. Therefore tf-idf will help us ignore terms that we might consider to be non-meaningful, given context of the data.

Our keyword extraction method is linear regression. To this end, we model each document in our corpus as a data point in j-dimensional space, where j is the number of unique terms encountered in our corpus. Linear regression helps us to find a j-dimensional vector that fits our corpus most effectively. The i'th term of this vector corresponds to the i'th term encountered in our corpus, and the value of the i'th element of this vector is a predictive weight assigned to that word which tells us the significance of that word to this corpus. Of course, significance is relative, and determined by how you model your data (whether it be by co-occurrence, or another metric). The keywords we present in our visualization correspond to words with the highest weights. We also impose a sparsity constraint in our linear regression model, where we try to set many of the elements in this resulting vector to be zero (this allows us to minimize extracted keywords that are only marginally relevant).

One limitation of our system is that it deals better with 'factual' data that it does with 'subjective' or 'opinionated' data. For example, the query "economy" on a decade-long news dataset not only successfully showed relevant terms like "banking", "government", and "regulations", but also showed a shift towards words like "slowing", "unemployment", and "crisis" near 2007. Empirically oriented, descriptive documents tend to work well with co-occurrence because they have relatively simple language and two terms that appear together frequently are thus more likely to be correlated.

On the other hand, analyzing Facebook comments proves to be a more difficult problem: comments are not written with the empirical conciseness or clarity found in empirically oriented, descriptive documents like financial news; the number of unique terms encountered is much higher; clusters formed between two co-

---

[16] Zi, L. *Tao Te Ching.* trans. Gia-fu Feng and Jane English. (1972).
Plato. *Apology* 21d.
Plato. Meno 80d 1-3.

occurring become less statistically significant. Additionally, there are relationships constructed within a thread of comments that don't necessarily exist outside of the thread. For example, a Facebook user named Danny posted a comment that incited an emotive response from other people in the same thread triggering a string of replies that all mentioned Danny by name. Our system extracted "Danny" as an important keyword because his name co-occurred with our query term. Danny, however, is not representatively descriptive (unless he goes viral).

      To deal with a limitation such as this, researchers must possess an understanding of the context in which data is embedded, as context cannot be quantified (or, thus, incorporated into our algorithmic analysis); " (Barnes 2004). A conversation is especially difficult for our system to algorithmically derive meaning from as each post relates to the post before it. The displayed comments are sometimes responses to another post, but without the context of the original post, the comments provide little information. A researcher must therefore have an initial knowledge of the viral events that happened in a time-slice for the related comments to make sense.

      Another limitation of this system is that it tends to highlight statistically dominant trends, but is unable to derive more subtle relationships between concepts. For example, if there are disagreeing opinions where one opinion is more strongly reflected by the data than another, the smaller voices will be marginalized. A similar example can be extracted from our previous analysis of the Bible.[17]

---

[17] Barnesmoore, L., Donoso, J., Claiver, S., El Ghaoui, L. (In Press) Machine Learning Methodologies: Histories of Asembalge and Representations of Women in the Bible. International Journal of Ciritical Cultural Studies.

Query: Any: All: woman
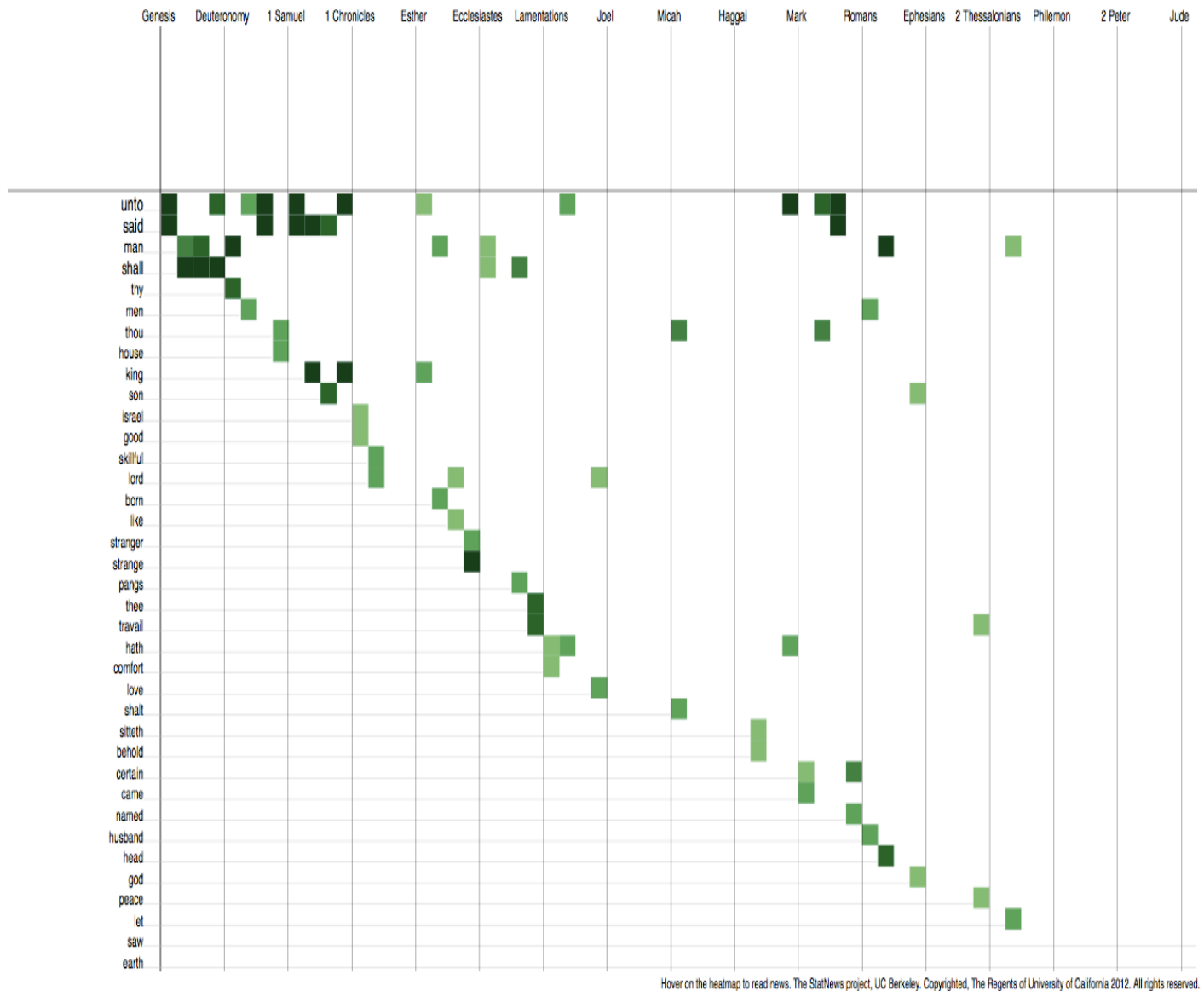
Statistically speaking, the story of Eve and Adam does not register in our algorithmic analysis of the term "woman". That being said, it is clear that this story is one of the most fundamental for understanding the relationship between the Bible and normative conceptions of women in Abrahamic societies. In these limitations the boundaries of the potential of quantification, especially as imposed by issues of meaning and symbolic representation, are clearly illustrated. Social Media comments, like symbolic religious texts, often rely on a few words to convey symbolic meaning. Symbolic meaning, even in a single moment (let alone in its evolution), cannot be quantified due to issues of dimensional incommensurability.[18] As a result, algorithmic attempts to analyze the meaning of large data text corpora in the social media sphere are often insufficient in of the dimensionally incommensurable relationship of symbolic meaning and quantification.

Netvizz

Netvizz is "a data collection and extraction application that allows researchers to export data in standard file formats from different sections of the Facebook social networking service."[19] Netvizz is a Facebook Application that programmatically makes requests to Facebook's servers using their API interface to retrieve

---

[18] See Ouspensky, P.D. *Tertium Organum* for a discussion of dimensional incommensurability.
[19] Rieder, B. (2013). Studying Facebook via data extraction: the Netvizz application. In Proceedings of the 5th Annual ACM Web Science Conference. New York: ACM, 346-355. http://rieder.polsys.net/files/rieder_websci.pdf
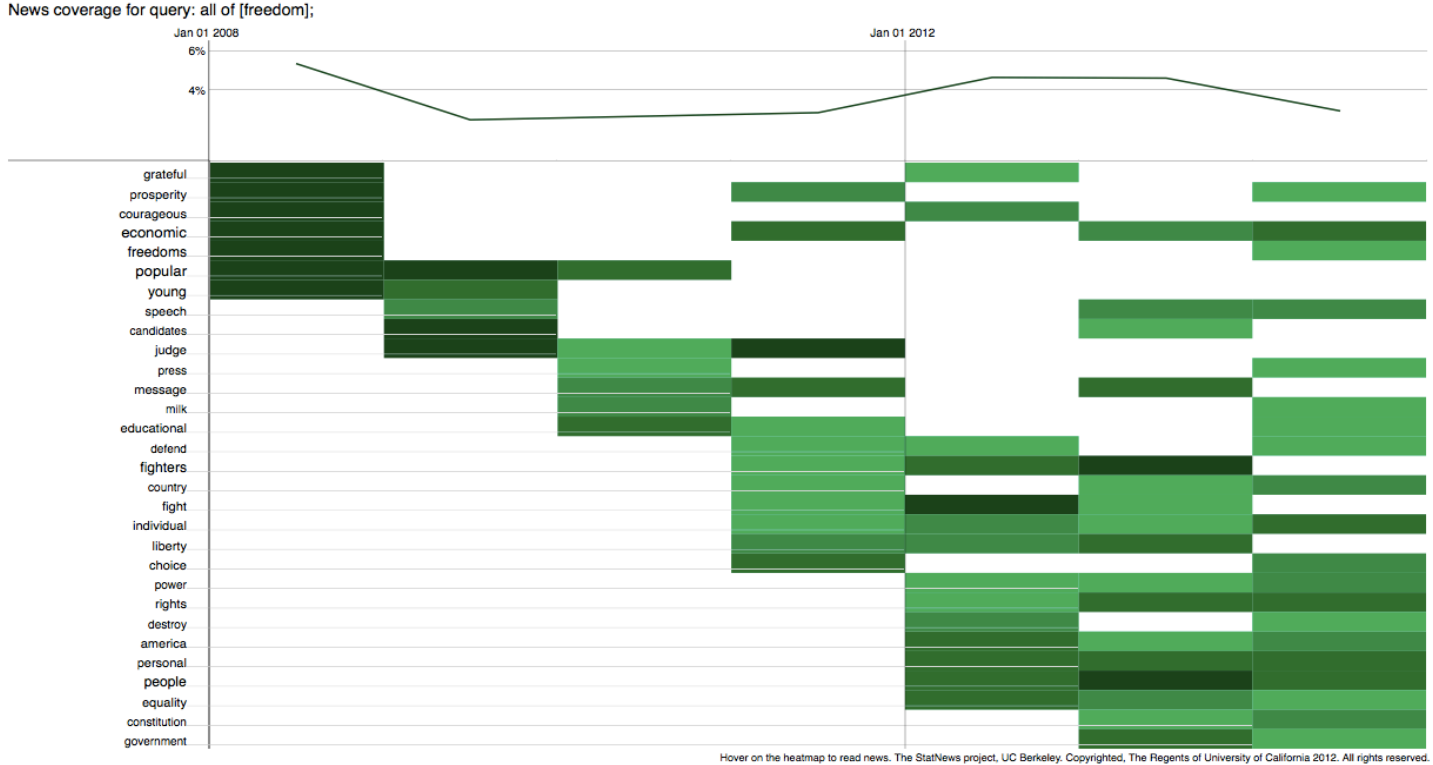
data. Because of the sensitive nature of social network data in the context of liberal democratic norms, Facebook requires authentication for people who use the Facebook API interface and only sends data through the API as a function of a particular user's Facebook network. For example, if you are not a part of a Facebook group on the main website, you are not allowed to see the posts and comments within that group. To programmatically tell Facebook which user you are, your Facebook account provides an authorization key and secret token, which are unique to your account. As an application built on top of Facebook, Netvizz can easily retrieve this authorization data by accessing your permissions allowing you to retrieve data sets drawn from your existing Facebook network with very little technical effort. The Facebook API imposes very few restrictions on the *amount* of data collected, so, while the Netvizz is unable to draw on Facebook data as a whole, the tool is especially useful for collecting large amounts of user data over a long timespan from within your Facebook network. In other words, you cannot collect all of Facebook's data in a single moment to analyze the overall narrative of Facebook data, but you can slice all of a group's data to analyze the overall narrative of the group through time.

Through integrating Netvizz and Facebook we ameliorate the cost of collecting data and writing code associated with software based language analysis approaches. Through integrating our Netvizz tool with the HoAM, we problematize narratives in the big data literature that argue "the era of big data challenges the way we live and interact with the world. Most strikingly, society will need to shed some of its obsession for causality in exchange for simple correlations: not knowing 'why' but only 'what'. This overturns centuries of established practices and challenges our most basic understanding of how to make decisions and comprehend reality."[20] The authors don't seem to provide all that much of an answer to why they are positing big data as signaling death of the question why, but as they deem such questions irrelevant in the big data era this is not all that surprising… In analyzing the meaning of language in relationship to publics socialized by language, we cannot forget questions of why as they articulate the essential meaning of the what. In other words, we at the same time open up the fields to whom the death of the question why has been attributed while attempting to revive the essential importance of the question why in understanding the essence of what therein. (Look back into the quoted text on the death of the question why to highlight the linkages between that paradigm and big data).
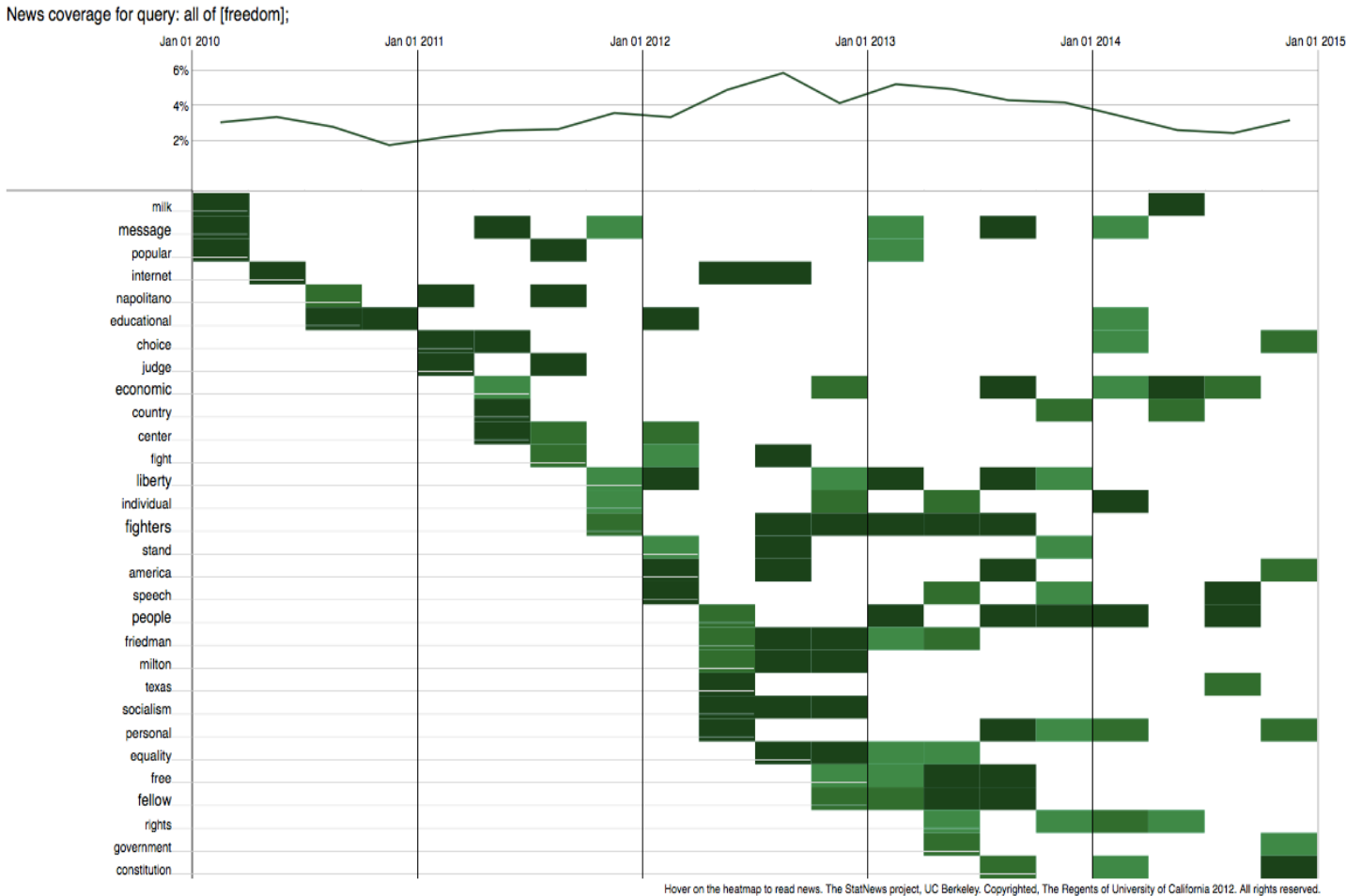
---

[20] Mayer-Schönberger, V. and Kenneth C. *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt, 2013., p. 6-7

Libertarian Facebook Narratives
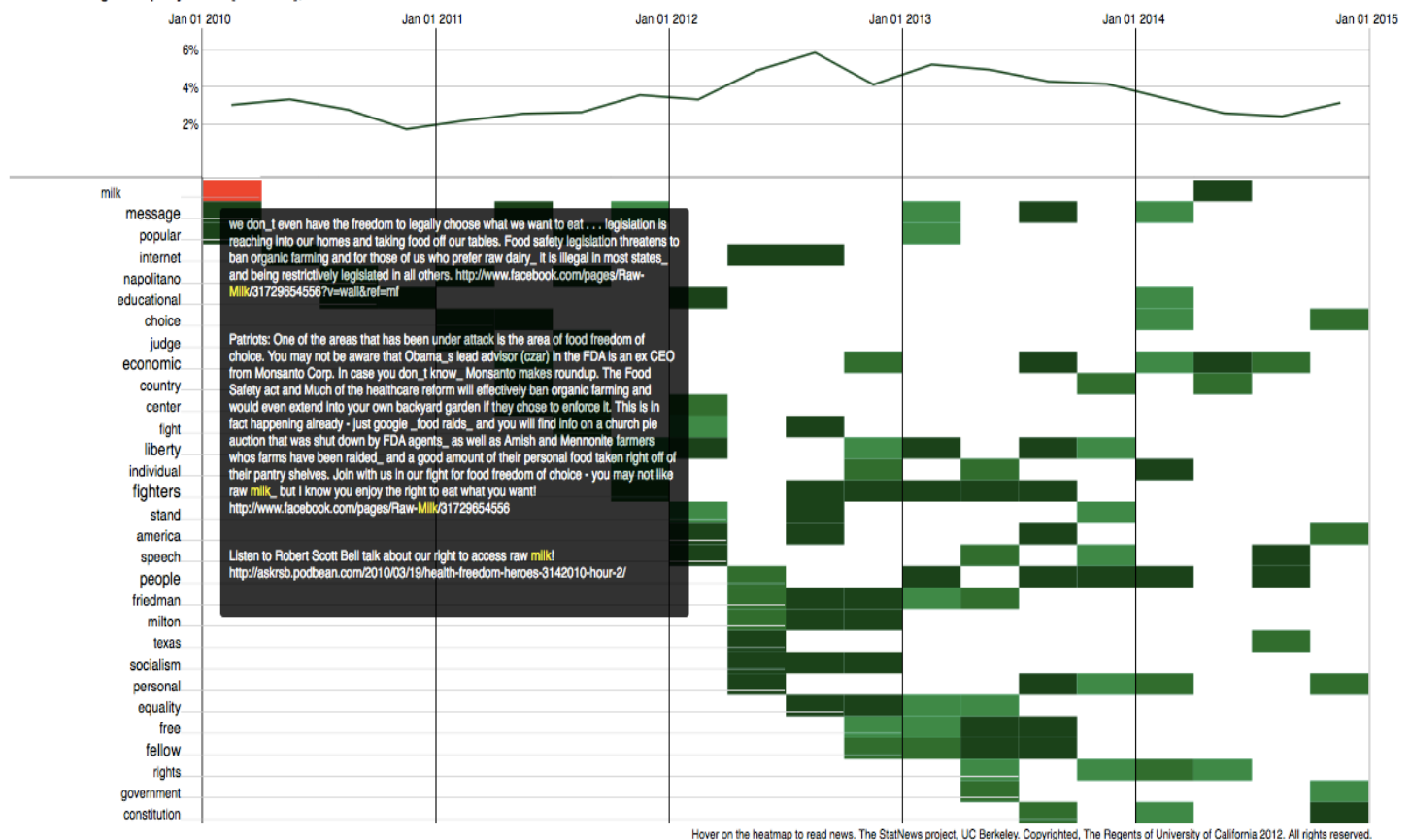
Freedom: 2008-2015 (Yearly)

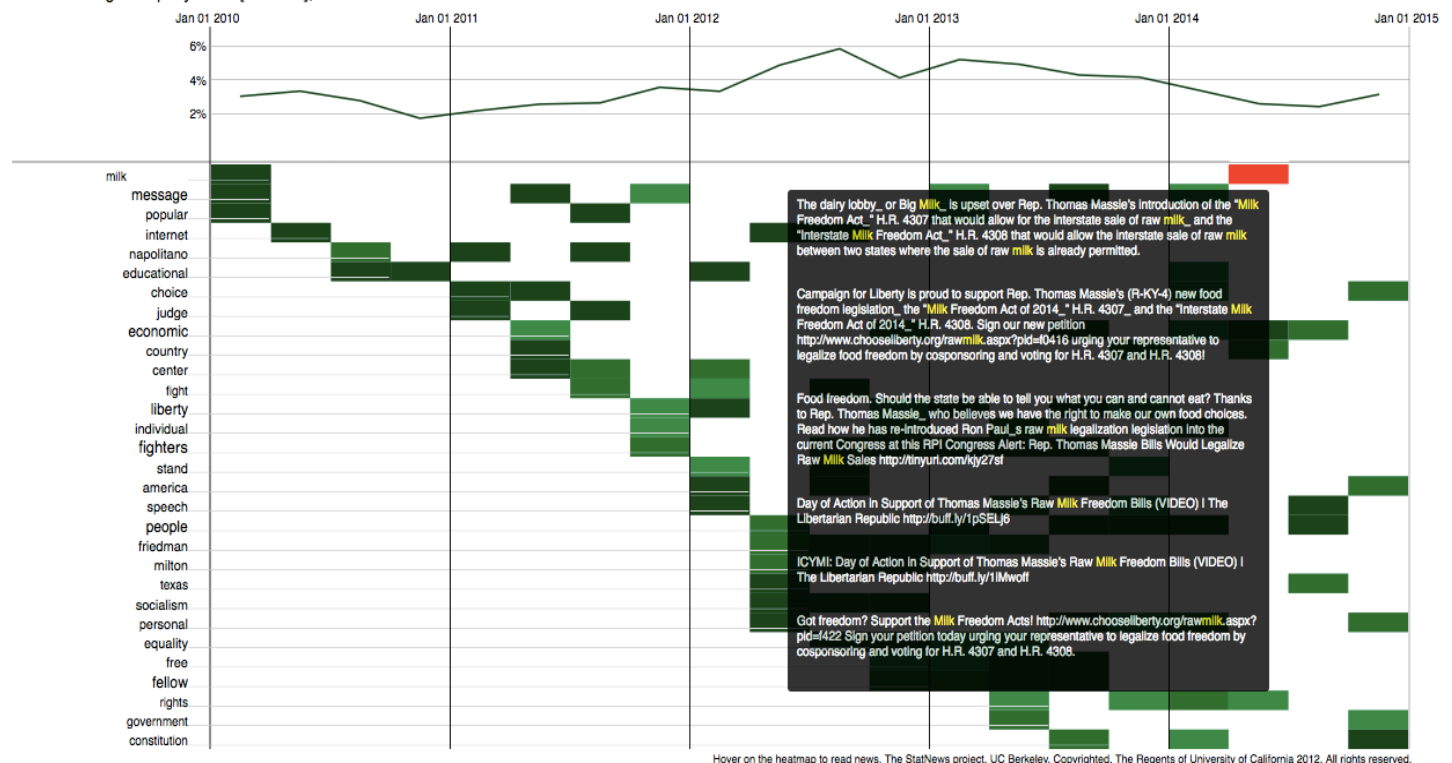News coverage for query: all of [freedom];

2010-2015 (Quarterly)

News coverage for query: all of [freedom];

Milk

News coverage for query: all of [freedom];

Jan 01 2010 Jan 01 2011 Jan 01 2012 Jan 01 2013 Jan 01 2014 Jan 01 2015

6%
4%
2%

milk
message
popular
internet
napolitano
educational
choice
judge
economic
country
center
fight
liberty
individual
fighters
stand
america
speech
people
friedman
milton
texas
socialism
personal
equality
free
fellow
rights
government
constitution

we don_t even have the freedom to legally choose what we want to eat . . . legislation is reaching into our homes and taking food off our tables. Food safety legislation threatens to ban organic farming and for those of us who prefer raw dairy_ it is illegal in most states_ and being restrictively legislated in all others. http://www.facebook.com/pages/Raw-Milk/31729456556?v=wall&ref=mf

Patriots: One of the areas that has been under attack is the area of food freedom of choice. You may not be aware that Obama_s lead advisor (czar) in the FDA is an ex CEO from Monsanto Corp. In case you don_t know_ Monsanto makes roundup. The Food Safety act and Much of the healthcare reform will effectively ban organic farming and would even extend into your own backyard garden if they chose to enforce it. This is in fact happening already - just google _food raids_ and you will find info on a church pie auction that was shut down by FDA agents_ as well as Amish and Mennonite farmers whos farms have been raided_ and a good amount of their personal food taken right off of their pantry shelves. Join with us in our fight for food freedom of choice - you may not like raw milk_ but I know you enjoy the right to eat what you want! http://www.facebook.com/pages/Raw-Milk/31729456556

Listen to Robert Scott Bell talk about our right to access raw milk! http://askrsb.podbean.com/2010/03/19/health-freedom-heroes-3142010-hour-2/

News coverage for query: all of [freedom];

Jan 01 2010 Jan 01 2011 Jan 01 2012 Jan 01 2013 Jan 01 2014 Jan 01 2015

6%
4%
2%

milk
message
popular
internet
napolitano
educational
choice
judge
economic
country
center
fight
liberty
individual
fighters
stand
america
speech
people
friedman
milton
texas
socialism
personal
equality
free
fellow
rights
government
constitution

The dairy lobby_ or Big Milk_ is upset over Rep. Thomas Massie's introduction of the "Milk Freedom Act_" H.R. 4307 that would allow for the interstate sale of raw milk_ and the "Interstate Milk Freedom Act_" H.R. 4308 that would allow the interstate sale of raw milk between two states where the sale of raw milk is already permitted.

Campaign for Liberty is proud to support Rep. Thomas Massie's (R-KY-4) new food freedom legislation_ the "Milk Freedom Act of 2014_" H.R. 4307_ and the "Interstate Milk Freedom Act of 2014_" H.R. 4308. Sign our new petition http://www.chooseliberty.org/rawmilk.aspx?pid=f0416 urging your representative to legalize food freedom by cosponsoring and voting for H.R. 4307 and H.R. 4308!

Food freedom. Should the state be able to tell you what you can and cannot eat? Thanks to Rep. Thomas Massie_ who believes we have the right to make our own food choices. Read how he has re-introduced Ron Paul_s raw milk legalization legislation into the current Congress at this RPI Congress Alert: Rep. Thomas Massie Bills Would Legalize Raw Milk Sales http://tinyurl.com/kjy27sf

Day of Action in Support of Thomas Massie's Raw Milk Freedom Bills (VIDEO) I The Libertarian Republic http://buff.ly/1pSELj6

ICYMI: Day of Action in Support of Thomas Massie's Raw Milk Freedom Bills (VIDEO) I The Libertarian Republic http://buff.ly/1lMwoff

Got freedom? Support the Milk Freedom Acts! http://www.chooseliberty.org/rawmilk.aspx?pid=f422 Sign your petition today urging your representative to legalize food freedom by cosponsoring and voting for H.R. 4307 and H.R. 4308.

The Cognitive environment of the US Libertarian Public

To frame our study of the relationship between ontological regimes and norms of thought behavior and being in the Libertarian public we begin by outlining the existing cognitive environment of the libertarian public (which colors the manifestation of ontological regime as norms of thought, behavior and being). We derive this outline

intuitively from the rational foundation (knowledge of aeons) developed in our daily interaction with the Facebook posts and user comments that form our analyzed dataset and our readings in the literature on Neoliberalism to allow our intuitive process to unfold this knowledge in a single movement of the mind.[21] American Libertarianism is first and foremost founded upon a neoliberal ontological regime (individualism, 'freedom' as capacity to do what one wants (especially with regard to market behavior), personal responsibility, market ethics, etc.). This cognitive environment of the libertarian public is also colored by ontological regimes associated with American Christianity (i.e. assumptions of evil human nature as a function of the fall, patriarchy, denigration of the other), American Nationalism (i.e. self-valorizing militarism, democratic veneer, racism, sexism, denigration of the other, etc.) and regional (place specific) socio-cultural norms. In short, the environment is typified by the banality of the assemblage of evils that lie at the heart of western oppression: egoic individualism, patriarchy, capitalism, Christian dogma, democratic veneer, reductive and oppressive definitions of freedom, etc. Again, as we noted in our methodological discussion, such attempts at generalization by necessity erase the actual contingency and complex difference of ontological regimes and their manifestation in a given public. That being said, the above generalization simply serves as a basic, though at a certain level inherently flawed (in generalizing), guide for our subsequent analysis.

Libertarian Freedom's Implicit Ontology in the Frame of Milk Rights
Before we begin our analysis of the narration of Freedom in the Libertarian Facebook Data Set we must problematize the banality of the term freedom through an intuitive discussion of traditional conceptions of freedom in the history of western (Hellenic) philosophy (as well as many other lineages of philosophy, notably those of China (Confucianism and Daoism) and the Abrahamic traditions). It seems that, in the modernist ontological framework (which we argue spans the continuum from positivist modernism to normative articulations of postmodernism in which neoliberalism is situated), Freedom is conceptualized as the ability to (if we may simplify) do, say, think, etc. what one wants. Taken to its logical conclusion, this conceptualization of freedom would necessitate that one actualize their desires as they see fit. While this conception of freedom may seem commonsensical in the context of modernist ontologies and neoliberal socialization, it is not the only model that has been posited through history. For many nodes of classical philosophy (again we generalize with the caveat that the general form we outline has manifested itself differently with regard to the context of manifestation), Freedom was associated with the transcendence of reflexive articulation of speech, behavior, thought, etc. as a function of the biological life functions (and the dimensional quality of passing time and physical space in general) through use of 'rational intuition' (Freedom is to transcend the conditioning by form associated with existence in passing time and physical space through purification of the self). From this lens, one is free to do the 'right thing' (to act as the virtuous subject would act) rather than free to do whatever they want; Freedom, then, rises as a function of the conception of reality one holds (one's relationship to truth) rather than the simple liberty to do as one wishes (one is, in this classical model, never free in the more simplistic modernist sense, as once one escapes from determinism by the dimensional quality of passing time and physical space they move into determinism by the 'right way' of harmonizing the uncreated and the created).

In analyzing the statistically dominant narration of 'Freedom' in the Libertarian Facebook Sphere we begin to observe the contours of the ontological regime implicit therein (as in the correlation of the term economic with freedom in the first visualization). Freedom, it seems clear, is defined within the ontological regime of neoliberalism and, more generally, modernism. In the neoliberal ontological regime implicit in our

[21] Ong, Aihwa. "Neoliberalism as a mobile technology." *Transactions of the Institute of British Geographers* 32, no. 1 (2007): 3-8.
Aihwa Ong and J Collier, *Global Assemblages* (Malden: Blackwell, 2005.)
Ong, Aihwa. *Neoliberalism as exception: Mutations in citizenship and sovereignty*. Duke University Press, 2006.
Harvey, David. *A brief history of neoliberalism*. Oxford University Press, 2005.
Rofel, Lisa. *Desiring China: Experiments in neoliberalism, sexuality, and public culture*. Duke University Press, 2007.
Duggan, Lisa. *The twilight of equality?: Neoliberalism, cultural politics, and the attack on democracy*. Beacon Press, 2012.
Stuart Hall, "The Toad in the Garden: Thatcherism among the Theorists," Cary Nelson & Lawrence Grossberg, *Marxism and the Interpretation of Culture* (University of Illinois Press, 1988).
Sut Jhally, "Image-Based Culture," Kelley Askew and Richard R. Wilk, the Anthropology of Media: a reader (Oxford: Blackwell Publishing, 2002) 327-336.
Sut Jhally, "Advertizing at the Edge of the Appocalypse," *SutJhally.com*, UMASS Amherst, <http://www.sutjhally.com/articles/advertisingattheed/>.
Stuart Ewen, *All consuming images: The politics of style in conteporary culture* (New York: Basic Books, 1988).
Brenner, Neil, Jamie Peck, and Nik Theodore. "Variegated neoliberalization: geographies, modalities, pathways." *Global networks* 10, no. 2 (2010): 182-222.
Lemke, Thomas. "'The birth of bio-politics': Michel Foucault's lecture at the Collège de France on neo-liberal governmentality." *Economy and society* 30, no. 2 (2001): 190-207.
Foucault, M. (2008). *The Birth of Biopolitics*, trans. Graham Burchell.

visualization, freedom is (if we may again simplify) the freedom to do what one wants (especially with regard to market behaviors). This conception of freedom is clearly implicit in the texts surrounding debates over milk rights extracted by our software. We analyze 'Milk Rights' both because it illustrates the limited nature of Libertarian-neoliberal conceptions of freedom and because it demonstrates the banality of this limited conception of freedom (debates over the nature of freedom give way to debates over absurd manifestations of a very limited conception of freedom due to the banality of the normative understanding of the nature of freedom).

"We don't even have the freedom to legally choose what we want to eat . . . legislation is reaching into our homes and taking food off our tables."

"One of the areas that has been under attack is the area of food freedom of choice. You may not be aware that Obama's lead advisor (czar) in the FDA is an ex CEO from Monsanto Corp. In case you don't know, Monsanto makes roundup. The Food Safety act and Much of the healthcare reform will effectively ban organic farming and would even extend into your own backyard garden if they chose to enforce it. This is in fact happening already - just google 'food raids' and you will find info on a church pie auction that was shut down by FDA agents, as well as Amish and Mennonite farmers whos farms have been raided, and a good amount of their personal food taken right off of their pantry shelves. Join with us in our fight for food freedom of choice - you may not like raw milk, but I know you enjoy the right to eat what you want!"

"The dairy lobby, or Big Milk, is upset over Rep. Thomas Massie's introduction of the "Milk Freedom Act_" H.R. 4307 that would allow for the interstate sale of raw milk, and the "Interstate Milk Freedom Act." H.R. 4308 that would allow the interstate sale of raw milk between two states where the sale of raw milk is already permitted."

"Campaign for Liberty is proud to support Rep. Thomas Massie's (R-KY-4) new food freedom legislation, the "Milk Freedom Act of 2014," H.R. 4307 and the "Interstate Milk Freedom Act of 2014," H.R. 4308. Sign our new petition http://www.chooseliberty.org/rawmilk.aspx?pid=f0416 urging your representative to legalize food freedom by cosponsoring and voting for H.R. 4307 and H.R. 4308!"

"Food freedom. Should the state be able to tell you what you can and cannot eat? Thanks to Rep. Thomas Massie, who believes we have the right to make our own food choices. Read how he has re-introduced Ron Paul's raw milk legalization legislation into the current Congress at this RPI Congress Alert: Rep. Thomas Massie Bills Would Legalize Raw Milk Sales http://tinyurl.com/kjy27sf."

There are two essential nodes of the libertarian ontological regime elucidated in these quotes. First, freedom is a function of our being able to do what we want, especially with regard to market behaviors. Second, the proper means for establishing and defending freedom is participation in the democratic system and legal reform. Our ability to move from terms that hold statistic relevance to the statistically dominant context of the terms use in the corpus (i.e. the above extracted texts) allows us to synthesize traditional statistical methods for language analysis associated with measuring co-occurrence (often associated with the work of authors like G.K. Zipf)[22] with modes of qualitative discursive analysis that necessitate analysis of actual text (rather than statistical trends in text) like the HoAM (in our example those associated with the work of authors like Foucault on discursive materials that orient themselves as guides to behavior).[23] In other words, our tool provides a mechanism for developing a representative, statistical foundation (upon an algorithmic subjectivity) for qualitative studies of large data text corpora. One might consider this foundation 'objective' in rising from a single algorithmic subjectivity. In other words, the methodology re-scales Zipf's hierarchal analysis of words through locating it within a multidimensional context amenable for qualitative analysis that requires knowledge of the contextual use of terms.

The power (and we would argue danger) of the Libertarian movement rises as a function of its synthesis of seemingly progressive political stances (for example, challenging the fascist relationship between the US Government and Monsanto) with a conservative, Modernist, Christian ontological regime. The conservative,

[22] Zipf, George Kingsley. "The Psycho-Biology of Language." (1935).
[23] Foucault, Michel. (2006) *History of Madness*, trans. Jonathan Murphy and Jean Khalfa. Random House LLC.
Foucault, Michel. (1977) *Discipline and punish: The birth of the prison*. Random House LLC.
Foucault, Michel. (1978) "The history of sexuality, volume 1: An introduction, Vol. 1. *New York: Vintage*.
Foucault, Michel. (1990) *The history of sexuality: The use of pleasure*. Vol. 2. Random House LLC,.
Foucault, Michel. (1986) "The care of the self: The history of sexuality, Vol. 3." *New York: Pantheon*.

Modernist and Christian nodes of the movement are often left implicit, as the power of such banally received assumptions lies in the silence that surrounds our words. So, while the Libertarian movement on the one hand purports to challenge the status quo in the United States political arena, it, on the other hand, propagates (often unconsciously) the very ontological regime that forms the foundation of elite class power in the modernist era (illusory conceptions of freedom associated with capitalism and democracy in this study's example) as well as many much older disciplinary ontological assumptions associated with the Hellenic and Abrahamic philosophy (the valorization and normalization of hierarchical social relationships implicit in the call for democratic, legal reform). The progressive veneer of the Libertarian movement grants force to the silent propagation of its oppressive and conservative ontological regime, as this ontological regime is akin to (and in the context of neoliberal conceptions of freedom almost identical to) the vast majority of the official (democrat and republican…) political spectrum in the United States (which again, as the silent propagation in the Libertarian movement, renders the ontological regime banal). The hegemonic ontological regime (of which neoliberal freedom is but a node) of modernism gleans its power directly from its silent transmission, from its silent, ubiquitous banality in the context of seemingly insurmountable difference at the level of expression—the power of hegemonic ontological regimes exists in the silence that surrounds our words.

Thought, Behavior and Being
In terms of thought (and in this example perception of being), this narration of freedom (as well as the implicit neoliberal ontological regime therein) is likely to lead people in the libertarian public to think of the world within an epistemological framework that axiomatically presumes the agency of the individual (though to varying degrees as a function of the relationship between the state and the market) within the capitalist system (the system of wage slavery). With regard to the example of milk rights, the push for legal reform as a way of challenging the ban on raw milk encourages people to think of the legal system as the proper means for establishing and defending their 'freedom' (and thus to think in terms that strip them of emancipatory potential).

In terms of behavior, there are two likely outcomes: first, people are likely to behave in egotistical ways as egoic behavior is implicitly valorized by narration of freedom as the ability to do what one wants (especially with regard to market behaviors within a system, capitalism, which structurally promotes individualistic, survivalist behavior); second, people are unlikely to take actions necessary to emancipate themselves from the oppression of wage slavery because, thinking of themselves as free, they are unmotivated to do so. This second likely outcome is clearly illustrated in the discourse around milk rights. By posing legal reform as the acceptable behavioral mechanism for establishing and defending one's freedom, people are socialized to behave in a manner which has no emancipatory potential with regard to the existing system (amelioration of oppression through means that play within the rules of the system only work to sustain the system through preventing people from taking actions that have the potential to destabilize the system). The same can be said of the frame of democracy implicit in these narratives of legal reform; by framing the choice between political candidates in the decrepitly corrupt US democratic system as freedom, people are lead to believe they are free (and act as though they are free) when in reality they don't even have freedom of choice (let alone true freedom in its classical conception…). This proposed behavioral framework is also rather ironic in its seeming tension with the libertarian ethos of challenging government regulation (government regulation should be combatted through government regulation and democratic participation…).

Bibliography

Americans for Limited Government. Facebook. https://www.facebook.com/AmericansForLimitedGovernment

Barnesmoore, L., El Ghaoui, L. (2013) "English vs. Arabic: Differing Al Jazeera Frames," International Studies Association 54th Annual National Convention San Francisco: ISA, 1-19. http://files.isanet.org/ConferenceArchive/e1419ee1b6af407090d3eab494f84379.pdf

Barnesmoore, L., El Ghaoui, L. Pham, V. (2014) "Media Frames & Histories of Assemblage," International Studies Association 55th Annual National Convention. Toronto: ISA, 1-57. http://web.isanet.org/Web/Conferences/Toronto%202014/Archive/1d0fb7f3-ee28-443a-9535-e2fd1f95e1ff.pdf

Barnesmoore, L., Donoso, J., Claiver, S., El Ghaoui, L. (In Press) Machine Learning Methodologies: Histories of Asembalge and Representations of Women in the Bible. International Journal of Ciritical Cultural Studies.

Brenner, N., Peck, J., Theodore, N. (2010) "Variegated Neoliberalization: Geographies, Modalities, Pathways." *Global Networks* 10, no. 2.

Campaign for Liberty. Facebook Page. https://www.facebook.com/CampaignforLiberty

CATO Institute. Facebook Page. https://www.facebook.com/CatoInstitute

Clavier, S., Barnesmoore, L., El Ghaoui, L., and Li, G. C. (2012) "All the News That's Fit to Compare," International Studies Association 53rd Annual National Convention San Diego: ISA 1-37. http://files.isanet.org/ConferenceArchive/bb6556e4811e410c9c524cc152ae68a5.pdf

Deleuze, G. (1987) *Thousand Plateaus*, trans. Brian Masumi University of Minnesota Press.

Descartes, R. (2010) *Discourse on the Method of Rightly Conducting one's Reason and Seeking Truth in the Sciences*, (http://www.earlymoderntexts.com/pdfs/descartes1637.pdf)

Descartes, R. (2002) *Meditations on First Philosophy,* (http://www.vahidnab.com/med.pdf)

Descartes, R. (2010) *The Passions of the Soul*, (http://www.earlymoderntexts.com/pdfs/descartes1649.pdf)

Duggan, L. (2012) *The Twilight of Equality?: Neoliberalism, Cultural Politics, and the Attack on Democracy*. Beacon Press.

Ewen, S. *All Consuming Images: The Politics of style in Conteporary Culture* (New York: Basic Books, 1988).

Foucault, Michel. (1977) *Discipline and punish: The birth of the prison*. Random House LLC.

Foucault, Michel. (1978) "The history of sexuality, volume 1: An introduction, Vol. 1. *New York: Vintage*.

Foucault, Michel. (1990) *The history of sexuality: The use of pleasure*. Vol. 2. Random House LLC,.

Foucault, Michel. (1986) "The care of the self: The history of sexuality, Vol. 3." *New York: Pantheon*.

Foucault, Michel. (2006) *History of Madness*, trans. Jonathan Murphy and Jean Khalfa. Random House LLC.

Foucault, M. (2008). *The Birth of Biopolitics*, trans. Graham Burchell.

Gawalt, B., Jia, J., Miratrix, L., El Ghaoui, L., Yu, B. and Clavier, S. (2010) Discovering word associations in news media via feature selection and sparse classification. In *Proc. 11th SACM SIGMM International Conference on Multimedia Information Retrieval*.

Harvey, D. (2005) *A Brief History of Neoliberalism*. Oxford University Press.

Jhally, S. "Image-Based Culture," Kelley Askew and Richard R. Wilk, the Anthropology of Media: a reader (Oxford: Blackwell Publishing, 2002) 327-336.

Jhally, S. "Advertizing at the Edge of the Appocalypse," *SutJhally.com*, UMASS Amherst, <http://www.sutjhally.com/articles/advertisingattheed/>.

Kirkpatrick, M. (2010) "Twitter to Sell 50% of All Tweets for $360k/Year Through Gnip" Readwrite.com, 17 Nov. 2010. http://readwrite.com/2010/11/17/twitter_to_sell_50_of_all_tweets_for_360kyear_thro (https://gnip.com/sources/twitter/historical/)

Lemke, T. "'The Birth of Bio-Politics': Michel Foucault's Lecture at the Collège de France on Neo-liberal Governmentality." *Economy and society* 30, no. 2 (2001): 190-207.

Libertarian Party. Facebook Page. https://www.facebook.com/libertarians

Libertarianism.org. Facebook Page. https://www.facebook.com/Libertarianism.org

Mayer-Schönberger, V. and Kenneth C. *Big data: A Revolution that will Transform how we Live, Work, and Think*. Houghton Mifflin Harcourt, 2013.

Miratrix, L., Jinzhu, J., Yu, B., Gawalt, B., El Ghaoui, L., Barnesmoore, L., Clavier, S., (2014) "Concise Comparative Summaries (CCS) of Large Text Corpora with a Human Experiment." The Annals of Applied Statistics 8 (1), 499--529. doi:10.1214/13-AOAS698. http://projecteuclid.org/euclid.aoas/1396966296.

Ong, A. (2007) "Neoliberalism as a mobile technology." *Transactions of the Institute of British Geographers* 32, no. 1

Ong, A. and Collier, J. (2005) *Global Assemblages*. Malden: Blackwell.

Ong, A. *Neoliberalism as exception: Mutations in citizenship and sovereignty*. Duke University Press, 2006.

Ouspensky, P. D. (2007) *Tertium Organum*. Digireads. com Publishing. http://holybooks.lichtenbergpress.netdna-cdn.com/wp-content/uploads/Tertium-Organum-by-P-D-Ouspensky.pdf

Plato. *Apology*.

Plato. Meno.

Rieder, B. (2013). Studying Facebook via Data Extraction: the Netvizz Application. In Proceedings of the 5th Annual ACM Web Science Conference. New York: ACM, 346-355. http://rieder.polsys.net/files/rieder_websci.pdf

Ron Paul Institute for Peace and Prosperity. Facebook Page. https://www.facebook.com/RonPaulInstitute

Rofel, L. (2007) *Desiring China: Experiments in Neoliberalism, Sexuality, and Public Culture*. Duke University Press, 2007.

Spinoza, B. (2002) *Complete Works*. trans. Samuel Shirley. Hackett, p. 26.

The Libertarian Republic. Facebook Page. https://www.facebook.com/libertarianrepublic

Wall, M. (2014) "Big Data: Are You Ready for Blast-off?" *BBC News*, 3 Mar. 2014. http://www.bbc.com/news/business-26383058

Zi, L. *Tao Te Ching.* trans. Gia-fu Feng and Jane English. (1972).

Zipf, George Kingsley. "The Psycho-Biology of Language." (1935).