CrossMark

Measuring and Modeling the U.S. Regulatory Ecosystem

Michael J Bommarito II[1,2] ⓘ · **Daniel Martin Katz**[1,2]

Abstract Over the last 23 years, the U.S. Securities and Exchange Commission has required over 34,000 companies to file over 165,000 annual reports. These reports, the so-called "Form 10-Ks," contain a characterization of a company's financial performance and its risks, including the regulatory environment in which a company operates. In this paper, we analyze over 4.5 million references to U.S. Federal Acts and Agencies contained within these reports to measure the regulatory ecosystem, in which companies are organisms inhabiting a regulatory environment. While individuals across the political, economic, and academic world frequently refer to trends in this regulatory ecosystem, far less attention has been paid to supporting such claims with large-scale, longitudinal data. In this paper, in addition to positing a model of regulatory ecosystems, we document an increase in the regulatory energy per filing, i.e., a warming "temperature." We also find that the diversity of the regulatory ecosystem has been increasing over the past two decades. These findings support the claim that regulatory activity and complexity are increasing, and this framework contributes an important step towards improving academic and policy discussions around legal complexity and regulation.

Keywords Complex systems · Regulation · Law · Model · Ecosystem · Legal complexity

1 Introduction

Economies, like ecosystems, exhibit dynamic, complex behaviors resulting from the interaction of "organisms" inhabiting, altering, and being altered by their "environment." In the case of economies, organisms can be seen as companies, and environments can be seen, at least in part, as regulations. Just as changes in the environment like rising temperature can harm or help organisms, either broadly or for specific regions or organisms, so too can reg-

✉ Michael J Bommarito II
michael.bommarito@gmail.com

1 Illinois Tech - Chicago Kent College of Law, Chicago, IL, USA

2 CodeX - The Stanford Center for Legal Informatics, Stanford, CA, USA

ulation harm or help companies. Yet unlike studies of biological ecosystems, studies of the economy have thus far lacked a longitudinal, empirical measure of fundamental quantities like "temperature" or "diversity." In this paper, we attempt to bridge this gap, finding support for the common claim that regulatory activity and complexity has increased over the last 20 years.

First, we develop a simple theoretical model of regulatory "ecosystems" that allows us to conceptualize regulatory energy and diversity. We then proceed to measure these quantities over time through an analysis of filings made by companies registered under the Securities and Exchange Commission (SEC) in the United States. These filings, the so-called "Form 10-Ks" that are made each year by companies registered under the Securities and Securities Exchange Acts of 1933 and 1934, provide a broad overview of company performance and risks. In particular, 10-Ks highlight regulatory risks and uncertainty that companies face, allowing for the documentation of longitudinal and comprehensive data related to the regulatory ecosystem. In total, we analyze more than 20 years, 30,000 companies, and 160,000 10-K reports to identify more than 4.5 million references to U.S. Federal regulatory Acts and Agencies. Using these references, we generate a reproducible, quantitative, and longitudinal measurement of the energy, "temperature," and diversity of the U.S. regulatory "ecosystem." We find a clear increase for all of these measures, with double- to triple-digit growth over the last 20 years. We believe this framework and its ongoing application represent a principled approach to the quantification of regulatory ecosystems, and we hope that this research can drive better-grounded discussions of legal complexity and policy design in the modern world.

2 Data

10-K filings have been the focus of many academic studies in finance and accounting [1–4], law [5,6] and other adjacent fields [7–9]. Many of these studies have focused upon questions such as whether the reporting requirements are achieving their intended purpose and the extent to which markets react to the information disclosed in such findings. While these are certainly worthy questions in their own right, we believe these filings reveal important patterns that, *at scale* and *over time*, provide meaningful insight into a range of other scientific questions.

Companies that meet the requirements for 10-K reporting expend significant and increasing resources to prepare these documents, typically with the assistance of accounting firms and lawyers. Indeed, the Annual Audit Fee Survey [10] conducted by the Financial Executives Research Foundation reveals a mean and median 2015 expense of $1.8M and $522,205, respectively, across over 6000 filers. As required by law, the figures and statements contained within these reports are certified and attested to by both a company's officers and its independent auditors, and companies and their officers have strong incentives to faithfully and comprehensively report. On the contrary, the competition for capital ensures that organizations do not over-report risks relative to industry peers, as this may frighten away investors. Therefore, unlike other sources of information, these 10-K annual reports are generally more likely to convey a comprehensive and balanced description of the environment in which a company operates.

Form 10-K filings generally contain at least four parts and fifteen schedules, which collectively offer a wealth of useful information about registered companies. These parts include a characterization of a company's financial health, legal risks, and other systematic and idiosyncratic factors, such as the nature of the regulatory environment in which it operates. Some of these factors, such as tax credits, may be positive, but the majority of listed regulatory factors

are present as risks. While there are a number of specific requirements under the law, firms and industries are provided with some latitude regarding how to satisfy reporting requirements. In addition, as explored in [1] and [5], there have been some important changes in the reporting rules over time. That said, SEC form templates and accounting firm standards result in more similarity than difference across firms and across time.

Through this exercise of risk factor disclosure, companies typically describe various sources of regulatory risk, including the laws and administrative agencies that are most relevant to their respective businesses. Consider, for example, the 2015 10-K filing of Trans Energy, Inc., an oil and gas exploration company. Among other statutory and agency references, their filing references both the Migratory Bird Treaty Act of 1918 and Endangered Species Act of 1973:

> **The Endangered Species Act** ("ESA") was established to protect endangered and threatened species. Pursuant to that act, if a species is listed as threatened or endangered, restrictions may be imposed on activities adversely affecting that species' habitat. Similar protections are offered to migratory birds under the **Migratory Bird Treaty Act**. The Company conducts operations on oil and natural gas leases that have species, such as raptors, that are listed and species, such as sage grouse, that could be listed as threatened or endangered under the ESA.

Across the broader set of required company disclosures, 10-K filings are filled with references such as these; some, such as citations to Sects. 13 and 15(d) of the Securities Exchange Act of 1934, are boilerplate and required by the SEC's forms. Others, such as the 756 references to the Migratory Bird Treaty Act of 1918, or the 27 references to the Price-Anderson Nuclear Industries Indemnity Act of 1957 over the last 23 years, are not.

In the mid-1990s, the SEC introduced its Electronic Data Gathering, Analysis, and Retrieval ("EDGAR") system. Since then, nearly all registered company 10-K reports have been uploaded and made available on EDGAR, resulting in more than 160,000 10-K reports accessible online. We retrieve these 10-K reports and build a multi-stage pipeline that identifies and normalizes references to Acts and Agencies. References are first identified through standard natural language processing techniques; once a reference fragment is identified, it is then passed through a second stage of normalization. As one example, many filers reference the "Gramm Leach Bliley Financial Services Modernization Act of 1999;" however, they do not do so using its full name, as above. Instead, they frequently refer to it as "GLB," "Graham Leach Bliley," "Gramm Leach Bliley," or the "Financial Services Modernization" Act. In order to handle this variation, we built a mapping for over 600 potential Act references, relying on a combination of the US Code, Wikipedia, and manual review. This mapping is then combined with fuzzy-string matching techniques to correct for spelling mistakes such as "Graham Leach Bliley." The result is a high-precision and high-recall extraction of 401 unique Federal Acts and 133 Agencies across our 23-year dataset. In total, we identify more than 4.5 million Act and Agency references contained in 10-K reports over the past 23 years.

3 Theory of Regulatory Ecosystems

Modern ecology and environmental science rely heavily on the spatial modeling of ecosystems, allowing researchers to analyze the dynamics and interactions of environment and inhabitants [11,12] in open systems. In such models, environments are typically represented by two- or three-dimensional spaces that may be discretized into grids. Each "cell" of a grid

may contain or interact with resources or organisms like timber, infrared radiation, or cattle. These cells may also provide an area over which measures like averages or sums may be calculated. For example, since the 1970s, scientists have been using satellite instrumentation and representations of the earth's surface to model systems related to surface temperature, rainfall, land use, and soil hydrology [13,14].

We follow this spatial systems approach to model the regulatory environment. In the general case with M companies and N regulations, we can represent the regulatory ecosystem as M vectors in an N-dimensional space. We refer to each of these M vectors as a company or filing "profile." These profiles can encode only for the binary presence of regulatory exposure—a 0 or 1 in each dimension for each regulation. The profiles can also encode for the number of references in that regulatory dimension—0 or more for each regulation. At present, the regulatory space consists of $N = 401$ dimensions—one for each Act currently identified in our data.

Initially, in the ground state where no regulation exists, these M company vectors are all equal to the zero vector. As institutions "perform work" by expending "regulatory energy" to regulate companies, however, one or more of these M company vectors move, i.e., their profile vectors increase away from the zero vector. In the simplest case, where institutions regulate only one dimension i and all companies are homogeneous, then all M vectors move along dimension i proportional to regulatory work. This change in position away from the ground state in regulatory space captures an increase in total energy in the system. Subsequently, institutions may continue to perform work to regulate or de-regulate companies, resulting in increased or decreased energy along that dimension i.

In reality, institutions may simultaneously regulate and de-regulate along many dimensions. Furthermore, regulatory energy may not translate to equal motion in all cases. In some cases, non-conservative forces like lobbying "frictions" may reduce motion. In other cases, companies may exhibit different "mass" or other heterogeneity, translating equal regulatory energy into unequal regulatory "motion." Companies may also mutate, grow, or die over time, changing both the behavior of institutions and other companies as well as their own exposure to regulatory forces. In any case, the result is that companies may occupy varying positions in regulatory space over time. However, by measuring the total change in position from the ground state, we therefore obtain a measure of total energy in the regulatory system at a given time.

In the simplest case where institutions regulate homogeneous companies and all companies experience regulation equally, all M companies occupy the same position in regulatory space. However, as these assumptions are relaxed, then the positions of companies may vary across regulatory space. Measuring the "diversity" or degree of variation between companies allows us to understand how far away from the homogeneous case a regulatory system is. Just as in ecology and physics, the degree of diversity between companies has important implications for the efficiency and fragility of systems.

4 Methodology and Results

To formalize these concepts with notation, let a 10-K profile $\mathbf{p}(a)$ in $\mathbb{Z}^N$ have element $p_i(a)$ equal to the number of references to Act i for a company a's annual filing. This vector $\mathbf{p}$ can then be normalized or projected. For example, we can project $\mathbf{p}$ from the number of references to a "bitstring" vector $\mathbf{b}(a)$, whose element $b_i(a)$ is equal to 1 if a company a's filing mentions Act i at least once, else 0. Alternatively, we can aggregate or normalize these filings by viewing them as a time-indexed collection. Let $F(t)$ be the matrix whose rows
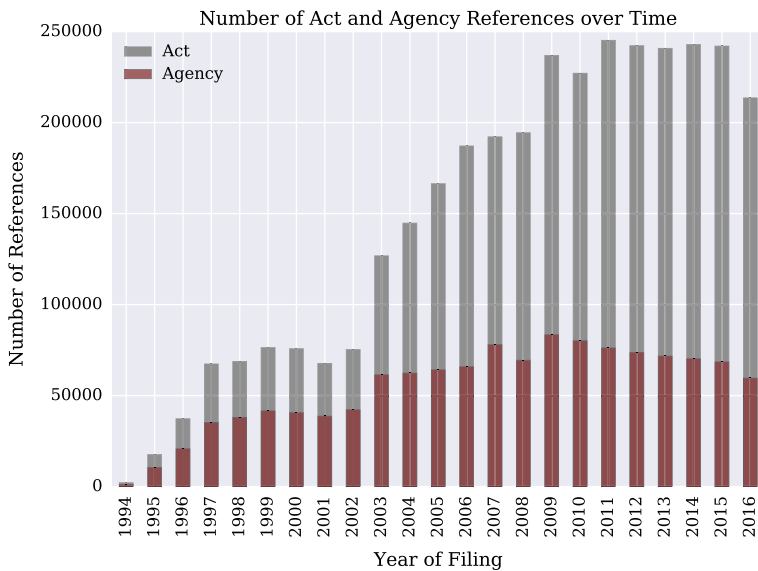
Fig. 1 Total number of Act references per filing over time

correspond to the set of all **p** or **b** vectors filed within a given year t. Then the grand sum of $F(t)$ is the total number of references, the jth column sum of $F(t)$ is the number of references to the jth Act, and the ith row sum is the number of references per firm in year t. It is possible to normalize the number of references per Act to a rate of reference per filing by dividing the column sums of $F(t)$ by the number of rows $m(t)$, which we call $\mathbf{r}_j(t)$ for jth Act in year t.

We then use these representations to measure the total energy, temperature, and diversity of the regulatory ecosystem as follows. First, we measure the total energy of the regulatory ecosystem using p-vectors as:

$$E(t) = \sum_i \sum_j F_{i,j}(t) \tag{1}$$

Figure 1 shows that the total energy, as measured by number of references to Acts per year, has increased substantially in the last 23 years. In 1996, there were just over 40,000 references to Acts in the nearly 5000 filings that year; by 2006, these numbers had more than quadrupled to nearly 200,000 references in just over 9000 filings; and, through three quarters of 2016, these numbers have again increased to an annualized rate of over 300,000 references.

Total energy alone can be misleading with respect to policy interpretations, however, as there are a number of reasons why energy may change without relation to regulatory exposure or "burden." For example, (i) the economy may grow or shrink in real or nominal terms, increasing or decreasing the total number of companies or companies meeting registration requirement, (ii) the SEC rules governing registration or filings may change, increasing or decreasing the number of companies or references, or (iii) *c.p.*, the relative incentives to incorporate or take on shareholders may change, increasing or decreasing the number of companies registered. These factors do not necessarily imply more or less regulation as experienced by individual companies, although they may be viewed as endogenous to some policy questions.
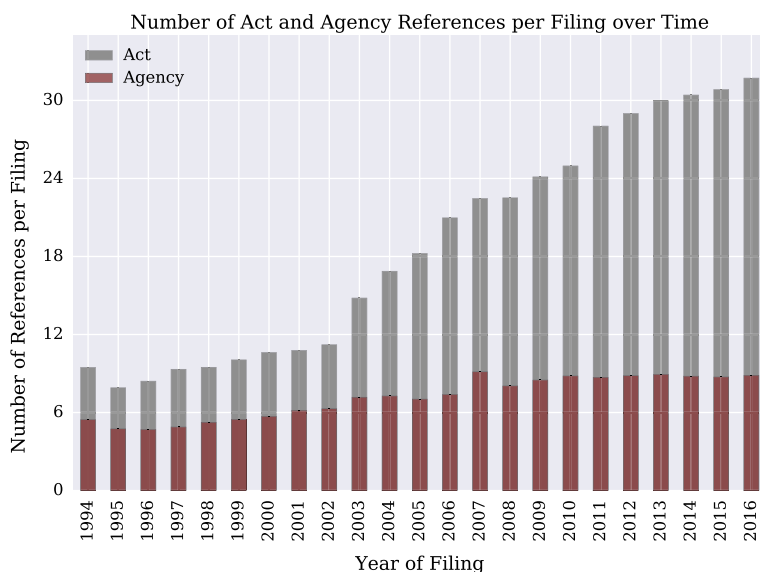
Fig. 2 Average number of Act references per filing over time

Table 1 Summary of energy and temperature measures over time

Year	$E(t)$	$T(t)$
1995	17,672	7.9
2000	75,851	10.6
2005	166,518	18.2
2010	227,210	25.0
2015	242,107	30.8

We can control for these factors by normalizing total energy to a "temperature," taking into account the number of filings per year as an analogy to area or volume. While this conception simplifies the traditional distinction between types of energy, we select it for its ecological analogy. To calculate, we take the average rate of references per filing, "temperature," $T(t)$ as follows:

$$T(t) = \frac{E(t)}{m(t)} \tag{2}$$

Figure 2 shows that, over the last 23 years, $T(t)$ has been monotonically increasing. While the rate of reference, like the total energy in Fig. 1 above, clearly shows the effect of the Sarbanes-Oxley changes in 2003, this trend remains unbroken both well before and well after. In 1996, the average number of references per filing was 8.4; by 2006, it had more than doubled to 20.9; and by 2016, the rate had increased again by more than 50% to 31.7 Act references per filing. Even if the amount of energy or cost does not scale linearly per filing with the number of references, the monotonic, 237% increase in $T(t)$ clearly demonstrates an increasing regulatory temperature.

Table 1 below summarizes the data from Figs. 1 and 2 above.

Finally, we may ask—is temperature or energy changing in concert with diversity, or is the change in temperature concentrated along a single dimension of regulation? For example,
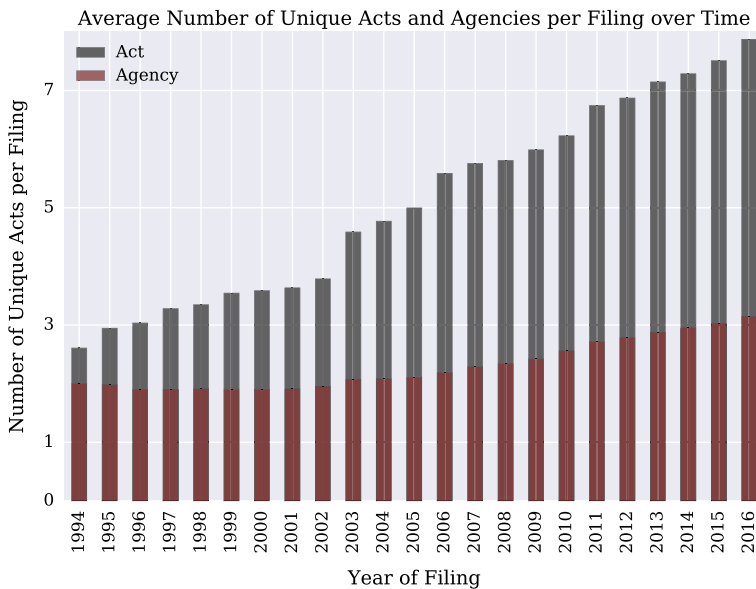
Fig. 3 Average number of unique Acts and Agencies per filing over time

changes in energy or temperature can represent more or less reliance on the same Act, e.g., the Securities Exchange Act of 1934; in this case, the number of unique Acts referenced is not changing, but the regulatory exposure per Act is. Alternatively, the total number of unique Acts referenced could be increasing or decreasing; for example, in 2003, most registered companies added references to the recently enacted Sarbanes-Oxley Act, which had not previously been referenced. Changes such as these represent an increase or decrease in the number of dimensions or diversity of regulatory exposure, but not necessarily the intensity of each exposure.

Using our notation above, we evaluate the diversity question by calculating two measures. First, we calculate the number of unique Acts per filing through the sum of b vectors above. Then, we calculate the average number of unique Acts per filing, across all companies in a given year; this is $\frac{1}{m} \sum_j \sum_i b_i^j$, where b_i^j is the bit corresponding to whether the ith Act was referenced in the jth company filing and m is the number of companies per year.

Figure 3 shows that, over the last 23 years, the diversity of Acts referenced has increased jointly with temperature. Like Figs. 1 and 2 above, the time series exhibits a jump following Sarbanes-Oxley; however, like Fig. 2, the time series also exhibits a monotonic increase over two decades, growing from 3.1 unique Acts per filing in 1996 to 5.6 unique Acts per filing in 2006 to 7.9 Acts per filing in 2016. This increase suggests that the increase in regulatory ecosystem temperature has been, at least in part, related to an increase in the number of dimensions along which institutions are regulating.

As an additional measure of diversity, we analyze each company's yearly regulatory "bitstring." As noted earlier, we calculate the 401-bit vector b for each company-year, where each bit corresponds to the presence of the 401 discrete Acts we identify. Although the regulatory space has 401 dimensions, the bitstring for a given filing is likely to be extremely sparse. For example, consider the 2012 10-K filed by the *Boeing Company*. Their filing features a bitstring with 12 non-zero elements, including Acts such as the Homeland Security Act, the Employee Retirement Income Security Act, the Patient Protection And Affordable
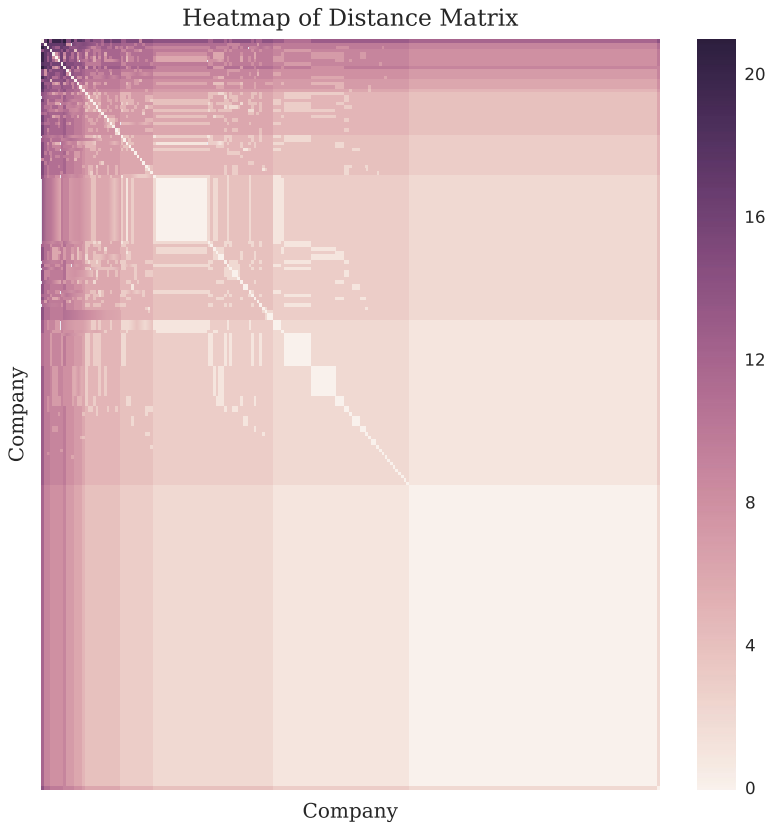
Heatmap of Distance Matrix

Fig. 4 Hamming distance heatmap for 1994 Act bitstrings

Care Act and the American Taxpayer Relief Act. Alternatively, the 2014 10-K of *Facebook Inc.* features 10 unique elements, including The Bank Secrecy Act, the U.S. Foreign Corrupt Practices Act, the USA Patriot Act, and the Credit CARD Act.

After applying this formalization to all companies for all years, we calculate the average pairwise Hamming distance [15] between all company bitstrings in a given year. Hamming distance is commonly used to evaluate the diversity of genomic [16–18] and other related data [19–21]. It can be interpreted as proportional to the average number of regulatory dimensions not in common between companies. More explicitly, the Hamming distance between two companies a and b in year t is:

$$d_{a,b}(t) = \sum_i \mathbf{p}(a) \oplus \mathbf{p}(b) \tag{3}$$

where $\oplus$ is the element-wise XOR operator. We can then write the average Hamming distance $\bar{d}(t)$ as the average over all combinations of a and b at t. Figure 4 visualizes the structure of the distance matrix D for all a, b as of 1994. The large block in the lower right corresponds primarily to special purpose vehicles like trusts or limited partnerships, and the overall structure corresponds to sectors and industries.

Figure 5, the average Hamming distance $\bar{d}(t)$ over time for Acts and Agencies, portrays mean-field distance between firms at scale, confirming an increasing diversity across the
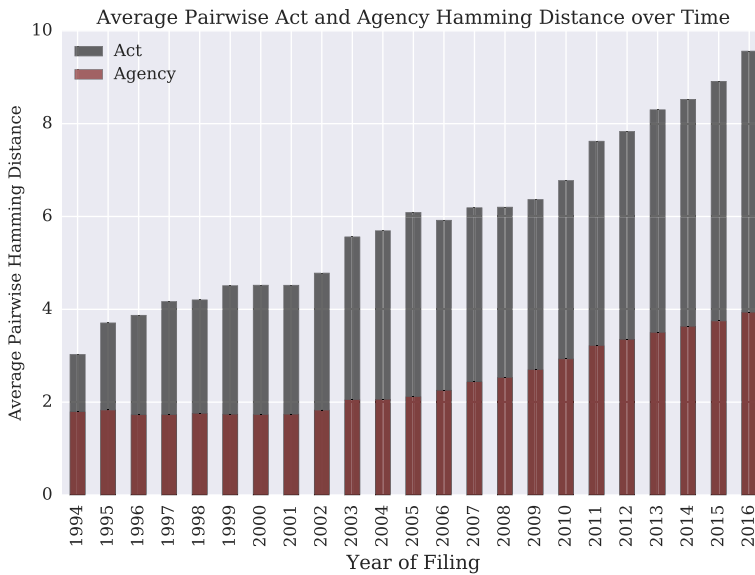
Fig. 5 Average pairwise Hamming distance between company filing regulatory bitstrings over time

Table 2 Summary of diversity measures over time

Year	Average unique acts	$\bar{d}(t)$
1995	2.9	3.7
2000	3.6	4.5
2005	5.0	6.1
2010	6.2	6.8
2015	7.5	8.9

global regulatory ecosystem. Over time, companies are subject, on average, to increasingly different requirements. While not monotonically increasing like the rate of reference and number of unique references above, the average distance increases 18 of 23 years in the sample. In 1996, two firms were separated on average by fewer than four regulatory Act "bits" or "genes"; by 2016, this number has increased to nearly 10.

Table 2 summarizes the data from Figs. 3 and 4 above.

5 Conclusion and Future Work

In this paper, we have presented the first large-scale, longitudinal characterization of the energy, "temperature," and diversity of the regulatory ecosystem as characterized by our spatial model. We have identified increasing regulatory exposure along an increasing number of dimensions, providing evidence in support of the claim that regulatory burden is increasing. Using a bitstring representation of firm regulatory exposure, we have confirmed that the aggregate Federal regulatory ecosystem is becoming more diverse over time, providing evidence in support of the claim that regulatory complexity is increasing. These conclusions

are based on more than 20 years, 30,000 companies, 160,000 10-K reports, and 4.5 million references contained in uniquely comprehensive and accurate 10-K reports.

In future work, we intend to expand upon these questions and connect to an extant research agenda, including the development of a more detailed theoretical framework, the modeling of agent-based or computational regulatory ecosystems, the categorization of regulatory "species" and "climates," and the integration of this analysis with our existing work on the complexity of other statutory, regulatory, and judicial systems [22,23].

Our work contributes to both the broader literature on legal complexity [24–27] and efforts to document the physical properties of legal systems as complex adaptive systems [28–33]. In addition, this paper is among a growing set of recent works applying tools of machine learning and natural language processing to better understand the behavior of various legal systems [34–36].

In sum, we believe that this framework for modeling and measurement will contribute to ongoing academic and policy discussions around legal complexity and policy design. The continued development of both global and specialized regulatory indices can provide for a principled, empirical basis of evaluation, standing in stark contrast to the vague generalizations that frequently guide current policy decisions.

References

1. Campbell, J.L., Chen, H., Dhaliwal, D.S., Lu, H.-M., Steele, L.B.: The information content of mandatory risk factor disclosures in corporate filings. Rev. Acc. Stud. **19**, 396–455 (2014)
2. You, H., Zhang, X.-J.: Financial reporting complexity and investor underreaction to 10-k information. Rev. Acc. Stud. **14**, 559–586 (2009)
3. Kravet, T., Muslu, V.: Textual risk disclosures and investors' risk perceptions. Rev. Acc. Stud. **18**, 1088–1122 (2013)
4. Li, F.: The information content of forward-looking statements in corporate filings-a naïve Bayesian machine learning approach. J. Acc. Res. **48**, 1049–1102 (2010)
5. Nelson, K.K., Pritchard, A.C.: Carrot or stick? The shift from voluntary to mandatory disclosure of risk factors. J. Empir. Leg. Stud. **13**, 266–297 (2016)
6. Doran, K.L., Quinn, E.L.: Climate change risk disclosure: a sector by sector analysis of sec 10-k filings from 1995–2008. N. C. J. Int. Law Commer. Regul. **34**, 101–147 (2009)
7. Bao, Y., Datta, A.: Simultaneously discovering and quantifying risk types from textual risk disclosures. Manage. Sci. **60**, 1371–1391 (2014)
8. Huang, K.-W., Li, Z.: A multilabel text classification algorithm for labeling risk factors in sec form 10-k. ACM Trans. Manage. Inf. Syst. (TMIS) **2**, 1–19 (2011)
9. Tsai, M.-F., Wang, C.-J.: On the risk prediction and analysis of soft information in finance reports. Eur. J. Oper. Res. **257**, 243–250 (2017)
10. F.E.R. Foundation, 2016 audit fee report, 2016
11. Tilman, D., Kareiva, P.M.: Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions, vol. 30. Princeton University Press (1997)
12. Levin, S.A.: Ecosystems and the biosphere as complex adaptive systems. Ecosystems **1**, 431–436 (1998)
13. Wilheit, T., Chang, A., Rao, M.V., Rodgers, E., Theon, J.S.: A satellite technique for quantitatively mapping rainfall rates over the oceans. J. Appl. Meteorol. **16**, 551–560 (1977)
14. Idso, S., Schmugge, T., Jackson, R., Reginato, R.: The utility of surface temperature measurements for the remote sensing of surface soil water status. J. Geophys. Res. **80**, 3044–3049 (1975)
15. Hamming, R.W.: Error detecting and error correcting codes. Bell Syst. Tech. J. **29**, 147–160 (1950)
16. Tully, D.C., Ogilvie, C.B., Batorsky, R.E., Bean, D.J., Power, K.A., Ghebremichael, M., Bedard, H.E., Gladden, A.D., Seese, A.M., Amero, M.A., et al.: Differences in the selection bottleneck between modes

of sexual transmission influence the genetic composition of the hiv-1 founder virus. PLoS Pathog. **12**, e1005619 (2016)

17. He, M.X., Petoukhov, S.V., Ricci, P.E.: Genetic code, hamming distance and stochastic matrices. Bull. Math. Biol. **66**, 1405–1421 (2004)
18. Mattiussi, C., Waibel, M., Floreano, D.: Measures of diversity for populations and distances between individuals with highly reorganizable genomes. Evol. Comput. **12**, 495–515 (2004)
19. Martin, E., Cao, E.: Euclidean chemical spaces from molecular fingerprints: hamming distance and hempel's ravens. J. Comput. Aided Mol. Des. **29**, 387–395 (2015)
20. Alfred, V.: Algorithms for Finding Patterns in Strings. Elsevier (2014)
21. Morrison, R.W., De Jong, K.A.: Measurement of population diversity. In: International Conference on Artificial Evolution (Evolution Artificielle), pp. 31–41. Springer
22. Bommarito, M.J., Katz, D.M.: A mathematical approach to the study of the United States code. Phys. A **389**, 4195–4200 (2010)
23. Katz, D.M., Bommarito II, M.J.: Measuring the complexity of the law: the United States code. Artif. Intell. Law **22**, 337–374 (2014)
24. Ruhl, J., Katz, D.M., Bommarito, M.J.: Harnessing legal complexity. Science **355**, 1377–1378 (2017)
25. Ruhl, J., Katz, D.M.: Measuring, monitoring, and managing legal complexity. Iowa Law Rev. **101**, 191–244 (2015)
26. Boulet, R., Mazzega, P., Bourcier, D.: Network analysis of the French environmental code. In: Ai Approaches to the Complexity of Legal Systems. Complex Systems, the Semantic Web, Ontologies, Argumentation, and Dialogue, pp. 39–53. Springer (2010)
27. Bourcier, D., Mazzega, P.: Toward measures of complexity in legal systems. In: Proceedings of the 11th International Conference on Artificial Intelligence and Law, pp. 211–215. ACM (2007)
28. Lee, E.D., Broedersz, C.P., Bialek, W.: Statistical mechanics of the US supreme court. J. Stat. Phys. 1–27 (2015)
29. Clark, T.S., Lauderdale, B.E.: The genealogy of law. Political Anal. **20**, 329–350 (2012)
30. Katz, D.M., Stafford, D.K.: Hustle and flow: a social network analysis of the American federal judiciary. Ohio State Law J. **71**, 457–507 (2010)
31. Smith, T.A.: The web of the law. San Diego L. Rev. **44**, 309–354 (2007)
32. Jones, G.T.: Dynamical jurisprudence: law as a complex system. Ga. St. UL Rev. **24**, 873–883 (2007)
33. Post, D.G., Eisen, M.B.: How long is the coastline of law? Thoughts on the fractal nature of legal systems. J. Leg. Stud. **29**, 545–584 (2000)
34. Aletras, N., Tsarapatsanis, D., Preoţiuc-Pietro, D., Lampos, V.: Predicting judicial decisions of the European court of human rights: a natural language processing perspective. PeerJ Comput. Sci. **2**, e93 (2016)
35. Katz, D.M., Bommarito II, M.J., Blackman, J.: A general approach for predicting the behavior of the supreme court of the United States. PLoS ONE **12**, e0174698 (2017)
36. McShane, B.B., Watson, O.P., Baker, T., Griffith, S.J.: Predicting securities fraud settlements and amounts: a hierarchical Bayesian model of federal securities class action lawsuits. J. Empir. Leg. Stud. **9**, 482–510 (2012)