Check List for Co-Host Prior to Meeting

- Make everyone co-hosts
- Test chat function
- Test breakout rooms(?)
- Confirm with tech that we want move across breakout rooms
- We don't want to give the tech permission to leave
- Find out what the AOM person should be doing (David Woods?)
- Can we set the countdown for dissolution of breakout rooms?
- Can we send global announcements across breakout rooms?
- We need to lay out the protocol

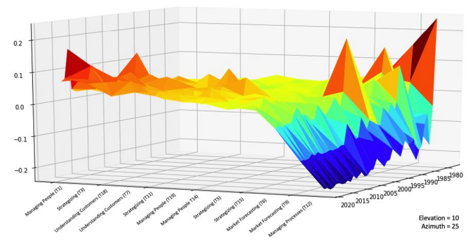
If you like, along with this deck, you can check out the following resources in advance….

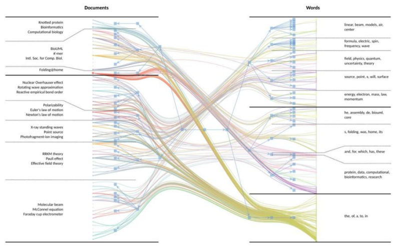
- Links for resources for the session
  - Hannigan, T. R., Haans, R. F., Vakili, K., Tchalian, H., Glaser, V. L., Wang, M. S., … & Jennings, P. D. (2019). Topic modeling in management research: Rendering new theory from textual data. *Academy of Management Annals, 13*(2), 586-632.
  - Schmiedel, T., Müller, O., & vom Brocke, J. (2019). Topic modeling as a strategy of inquiry in organizational research: A tutorial with an application example on organizational culture. *Organizational Research Methods, 22*(4), 941-968.
  - Other listed papers at the sites below.

- IDeaS general page: http://www.interpretivedatascience.com/

- GitHub: https://ideas-repo.github.io

- **If you examine this this deck in advance, we suggest that you refrain from looking at the exercise answers until after we do them collectively. ☺**





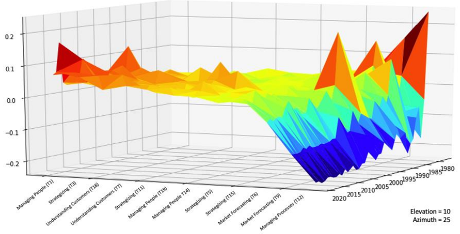
Topic Modeling Advances

Curating Corpora, Using Structured Models, and Theorizing with Visuals

ACADEMY OF Management
#AOM2020
80th Annual Meeting of the Academy of Management

Hannigan, Haans, Glaser, Tchalian, Valadao, Jennings
IDeaS
August 7, 2020

Goals of Session



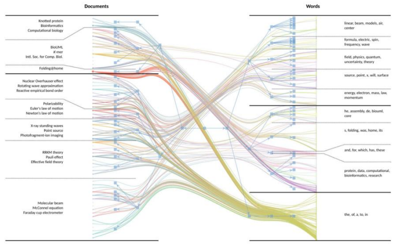
1. Build community

2. Extend your knowledge of advances in rendering with tmodeling.

Plan of the Session

Time (Mins)	Topic	Presenters
5-7	**Welcome, Goals, Plan, Protocols**	Dev (or Tim)
5-7	**Rendering with a Focus on Visuals**	Vern
15+5	**LDA & STM - Curation & Topic Modeling Methods for Visuals**	Richard & Rodrigo Informational Questions Only Please
15+15	**STM Exercise in Breakout Rooms (6) – Interactive (How to read output – 2 components)**	Set Up- Random Assignment + General Discussion
15+5	**hSBM & Hierarchical Models - Curation & Topic Modeling Methods for Visuals**	Tim & Hovig Informational Questions Only Please
15 + 15	**hSBM Exercise in Breakout Rooms (6) – Interactive (What to do with the outputs – 3 components)**	Set Up- Random Assignment + + General Discussion
10	**Back to Rendering Visuals, & Visualization Theory**	Dev
5	**Next IDeaS Workshop, Resources**	Dev (but all please chime in ☺)

Session Protocol



- Again, thanks for joining us; we look forward to your engagement.
- Please download the three key files on the AoM Web Page for this session (under the Files tab): this powerpoint (pdf form), two .xls exercise files.
- During the session, please raise your hand using the Zoom function and mute your mic until you're ready to ask questions.
- We'll use random assignment to breakout rooms of 5; if you leave a room, you'll have to request to be put back in manually.
- Sorry but no mid-point refresh, so please take the time when you have a moment (for standing, etc.).

How can we use topic modeling to generate new theoretical insights?

Topic Modeling Rendering in Theory-Building Spaces

Rendering corpora

Selecting
Trimming

Rendering topics

Applying algorithms
Fitting

Rendering theoretical artifacts

Creating
Building with

Visualization: A key step in the rendering process

Topic 32

Summary:

Words	*major*	*rebellion*	*job*	*event*	*state*	*report*	*case*	*crime*	*level*	*related*

Raw Topic Output:

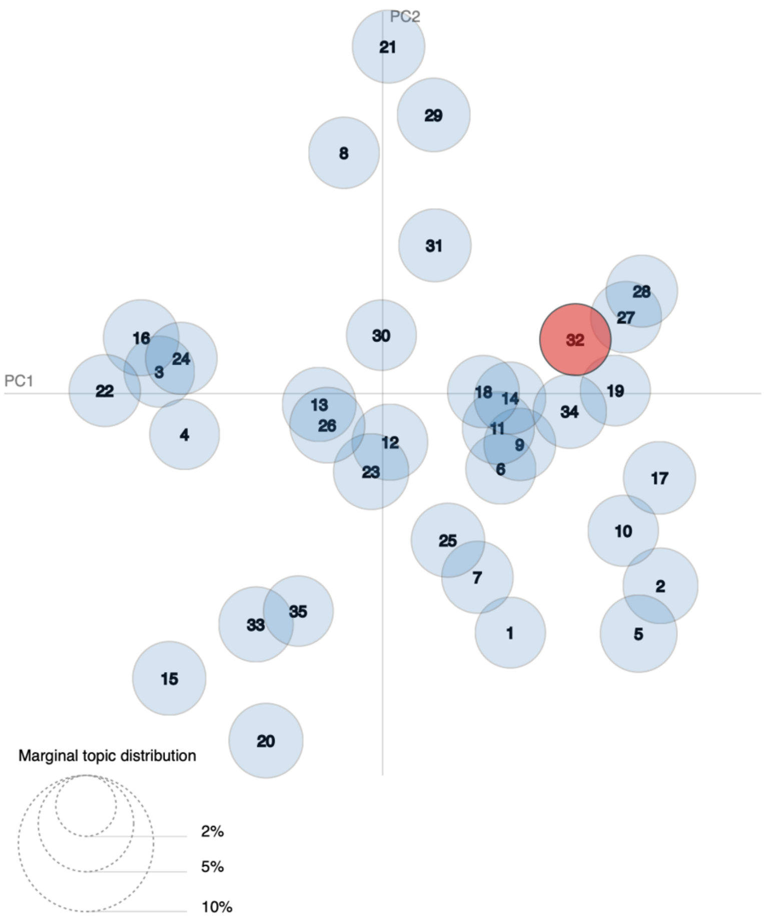
0.023*"major" + 0.019*"rebellion" + 0.018*"job" + 0.017*"event" + 0.015*"state" + 0.015*"report" + 0.014*"case" + 0.013*"crime" +
0.012*"level" + 0.010*"relate" + 0.010*"matter" + 0.010*"capture" + 0.010*"record" + 0.010*"law" + 0.009*"unrest"

Articles:

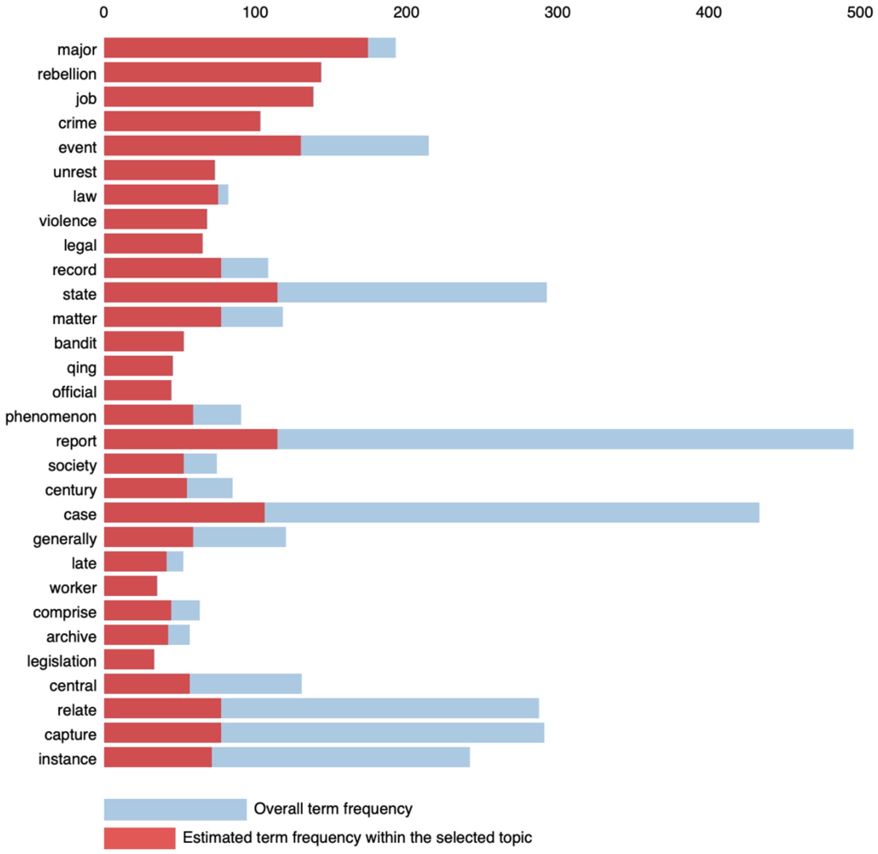
--weight: 0.57 title: Rebellion, crime and violence in Qing China, 1722-1911: A topic modeling approach - Miller, IM , 2013 POETICS
--weight: 0.16 title: Institutionalizing Big Data methods in social and political research - Ahonen, P , 2015 BIG DATA & SOCIETY
--weight: 0.09 title: Text Mining in Organizational Research - Kobayashi V.B., Mol S.T., Berkers H.A., Kismihók G., Den Hartog D.N. , 2018 Organizational Research Method
--weight: 0.07 title: Beyond Keywords: Tracking the Evolution of Conversational Clusters in Social Media
- Houghton J.P., Siegel M., Madnick S., Tounaka N., Nakamura K., Sugiyama T., Nakagawa D., Shirnen B. , 2017 Sociological Methods and Resea
--weight: 0.07 title: Introduction-Topic models: What they are and why they matter - Mohr, JW; Bogdanov, P , 2013 POETICS

Visualization: A key step in the rendering process

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 32 (2.8% of tokens)



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

The Importance of Visuals

- Theories of institutions, culture, relationality and neo-structuralism have all underscored the importance of visuals:
  - As artifacts (especially symbols) in cultures
  - As boundary objects in field relations
  - As representations of deeper structure
  - As rhetorical devices
  - As improved measures of extant concepts.

Two illustrations of visualization in advanced topic modeling techniques

1. Structural topic modeling (STM)
   - (Richard Haans and Rodrigo Valadao)

2. Hierarchical stochastic block modeling (hSBM)
   - (Tim Hannigan and Hovig Tchalian)

Moving Beyond LDA & Standard Topic Modeling – STM

Example / Exercise Part I

(Richard Haans & Rodrigo Valadao)

ACADEMY OF Management

#AOM2020

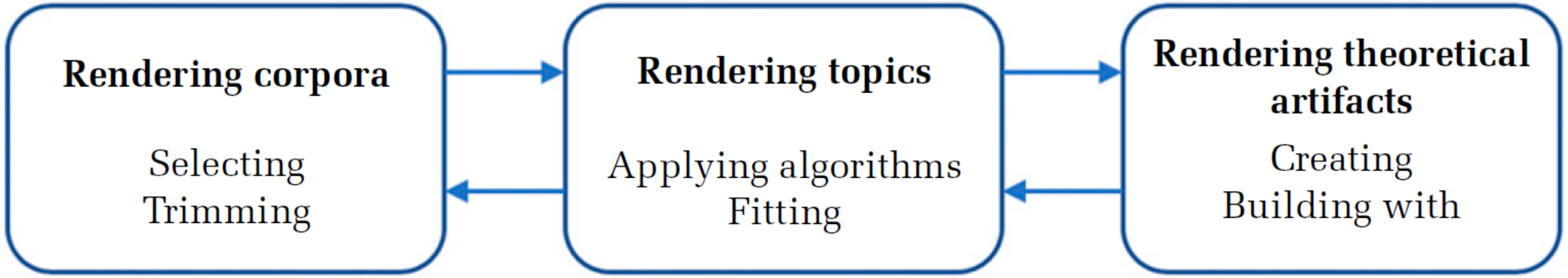
80th Annual Meeting of the Academy of Management

Hannigan, Haans, Glaser, Tchalian, Valadao, Jennings

IDeaS

August 7, 2020

STM and rendering



Metadata can more directly inform sample selection; cleaning of metadata more important ex ante.

Novel algorithms, model fit as determined not just by fit statistics but also now considering degrees of freedom (variables / topics ratios).

Novel theoretical artifacts such as covariate-topic effect estimates.

Basic LDA

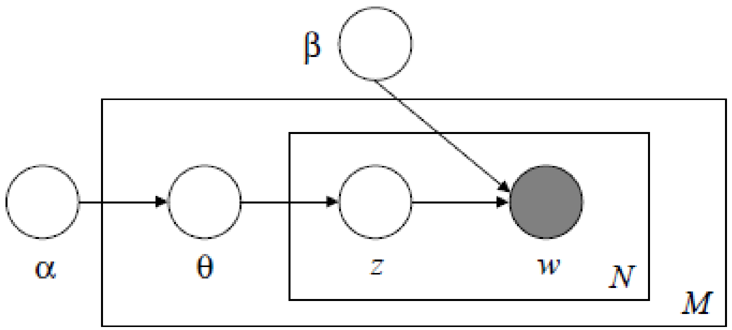


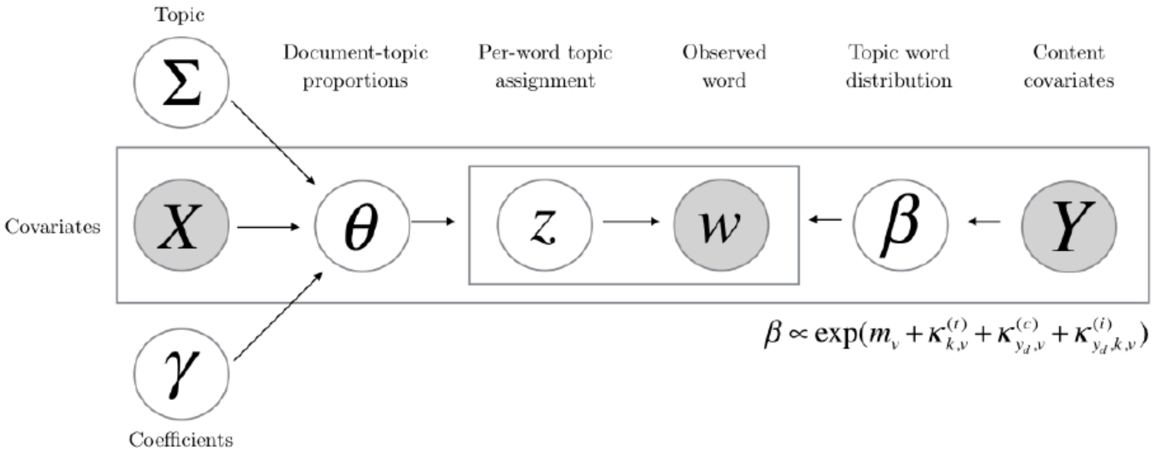
Figure 1: Graphical model representation of LDA. The boxes are "plates" representing replicates. The outer plate represents documents, while the inner plate represents the repeated choice of topics and words within a document.

(darker circle is what is observed; white circles what is learned)

The topic model learns **only** from the observed words co-occurrences in documents.

Assumption: identical generative processes behind texts in a corpora: documents are created based on drawing from a fixed set of topics—unchanging over time, independent of who generated the topics, etc.

Structural TM



A recent innovation is the ability to incorporate information from metadata into the estimation of the topic-word distribution or the document-topic proportions. Most common is the latter.

This enables understanding how e.g. characteristics of the document producer or contextual factors shape the extent to which topics are used in documents.

Canonical references

- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science, 58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103
- Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A Model of Text for Experimentation in the Social Sciences. *Journal of the American Statistical Association, 111*(515), 988–1003. https://doi.org/10.1080/01621459.2016.1141684
- Schmiedel, T., Müller, O., & vom Brocke, J. (2018). Topic Modeling as a Strategy of Inquiry in Organizational Research: A Tutorial With an Application Example on Organizational Culture. *Organizational Research Methods*, 109442811877385. https://doi.org/10.1177/1094428118773858

See: https://www.structuraltopicmodel.com/ for info on the method and applications.

In particular: the **'STM'** package in R opens many uses.

‘STM’ package

The STM package in R is attractive because essentially works from the basic topic modeling approach using packages like ‘topicmodels’.

See e.g. the materials from https://ideas-repo.github.io/workshops/

The metadata that was used *ex post* in those sessions can now be incorporated directly into the topic model.

Usual workflow (the practice)

1. Read the corpus; clean as usual using packages like '*tm*' (see 2017 session).
2. Turn corpus into document-term matrix.
3. Read the metadata; make sure sequence of documents is identical to the sequence in the corpus / dtm (sorting can be different).
4. Use the *readCorpus* function with "*type = c("slam")*"to convert the dtm to the STM format. Add the metadata (again; sorting has to be identical).
5. Use functions like *searchK* to identify best-fitting number of topics; *stm* to estimate the selected model; *labelTopics* to render topics; and *estimateEffect* and *plot* to interpret effects of covariates.

A Research Application & Exercise

EARLY MOMENTS OF INSTITUTIONAL CHANGE

Enable and constrain **possibilities** for institutional change.

(Hannigan & Casasnovas, 2020; Obstfeld et al., 2020; Thompson et al., 2018; Lounsbury and Glynn, 2019)

POSSIBILITIES

- Discursively constituted.
- Precursor to action.
- Interstices of identity positions → shaped by **category schemas**.

(Lounsbury and Glynn, 2019)

CATEGORY SCHEMAS

Analytically, by the time product categories coalesce in a market or industry, **we have already lost sight of the early moments of institutional change**.

MEANING INFRASTRUCTURE

- Product categories → cognitive infrastructure of markets *(Lounsbury and Rao, 2004)*.
- Meaning infrastructure → **building blocks of an underlying institutional meaning system**.

How to capture the *meaning infrastructure* that shapes the space of possibilities at the early moments of institutional change?

Methods

EMPIRICAL SETTING

The Emerging Field of Data Science: a case of an emergent *(Maguire et al., 2004)*, **interstitial issue field** *(Zietsma et al., 2017)*

DATA COLLECTION

- **Abductive approach:** 30 semi-structured interviews and 185 archival interviews with well-positioned actors (e.g., chief-data scientists, VPs of data science), conducted between 2017 and 2019.
- **Computational Approach:** Search on EBSCOhost Web for publications by **Harvard Business Review articles that contained the keywords "data" or "analy*" and issued between 1978 and 2018 (N = 3,005)**.

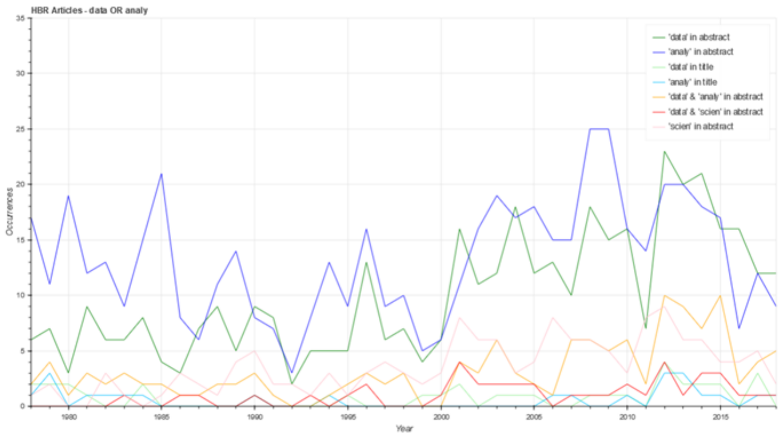
DATA ANALYSIS

- **Abductive approach:** close examination of the interviews to develop a rich understanding of the empirical context (Kaplan, 2015; see Hirsch & Lounsbury, 1997 and Lounsbury, 1998, 2001 for a similar methodological approach).
- **Computational approach:** **Structural Topic Modeling (STM) of the HBR articles** following a three-stages rendering process (see Hannigan et al., 2019).

ACADEMY OF Management
#AOM2020
80th Annual Meeting of the Academy of Management

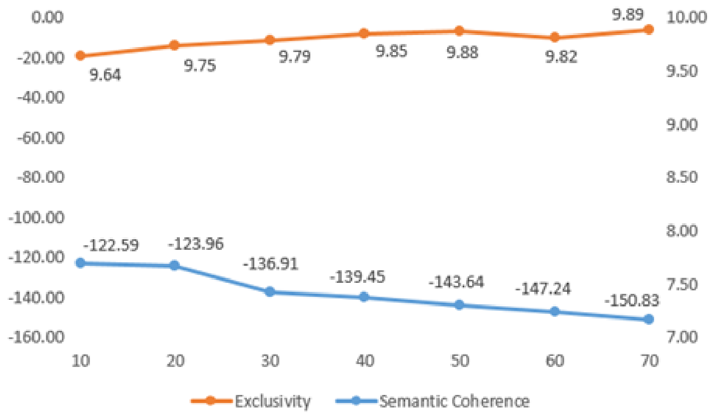
Rendering with STM (our demo)

RENDERING A CORPUS

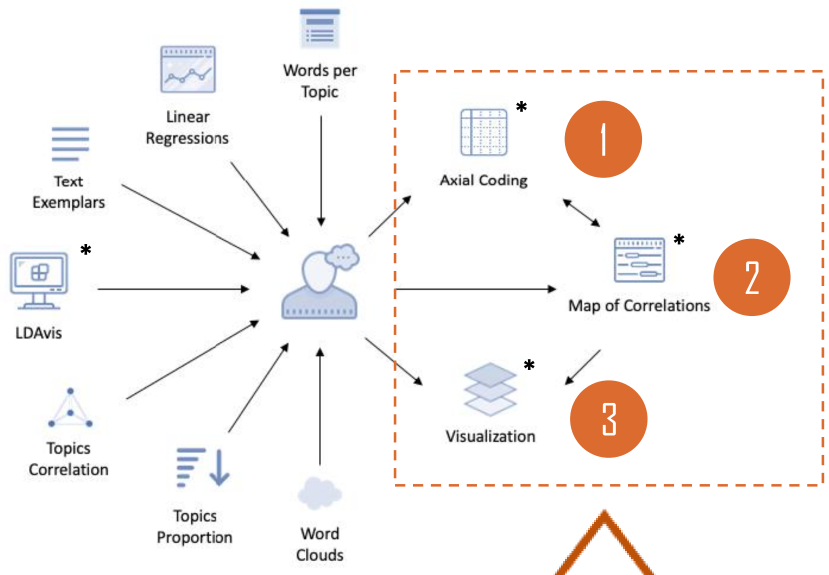


"Data" & "Analy*" in the abstract

RENDERING TOPICS



RENDERING THEORETICAL ARTIFACTS



* Not part of the *stm* package in R.

These 3 artifacts and additional instructions are going to be provided for discussion in the Break-Out session

ACADEMY OF Management
#AOM2020
80th Annual Meeting of the Academy of Management

Rodrigo Valadao, University of Alberta, August 7th 2020

Axial Coding

- Topics interpreted as elementary categories of meaning that constitute the meaning infrastructure of the emerging field of data science.
- STM x "Standard" Topic Modeling

1

Topic	Raw Words	First Order Codes	Second Order Codes	Aggregate Level	Ref.
1	Highest Prob: leader, execut, women, leadership, hire, talent, divers, cant, men, analyst FREX: women, men, femal, male, mentor, ect, divers, leadership, star, hire	Using analytics to compare leadership and to promote diversity (i.e. women, afro-americans, hispanics) in organizations	Promoting workplace diversity	Managing People	A
14	Highest Prob: perform, sale, firm, pay, forc, effect, number, rep, compens, among FREX: rep, salesperson, pay, incent, quota, micromarket, metric, perform, reward, fleet	Using data related to traing, incentives and performance to determine how to manage the sales force of an organization	Managing people based on analytics	Managing People	A
19	Highest Prob: employe, peopl, team, need, work, better, learn, oper, job, find FREX: island, team, profi, done, stori, baghai, invad, employe, head, knew	Discussing how the performance of a team can be misperceived and how other factors (i.e. external politics, management neglect) can have influence on the outcomes	Evaluating team performance	Managing People	A
3	Highest Prob: analyt, work, research, experi, insight, idea, project, program, knowledg, expert FREX: knowledg, scientif, human, social, insight, embrac, analyt, idea, collabor, expert	Explaining the concepts of scientific theory and experimentation, and how it influences the development of new technologies and products	Forstering research and innovation	Strategizing	B
5	Highest Prob: manag, busi, invest, may, review, strateg, capit, portfolio, asset, take FREX: review, busi, portfolio, invest, asset, ventur, loneli, extern, manag, owner	Using analytics to recommend investments in the long term and in large scale to investors, members of boards and leadership of asset owners	Allocating resources	Strategizing	B
11	Highest Prob: compani, will, make, decis, strategi, focus, chang, opportun, import, way FREX: decis, internet, structur, make, opportun, focus, will, strategi, shift, perspect	Assessing the impact of digitalization in strategic decisions of the organization and in the way they do business	Innovation shaping decisions	Strategizing	B
15	Highest Prob: develop, new, problem, model, need, can, improv, help, tool, complex FREX: problem, softwar, solut, complex, engin, involv, innov, develop, capabl, model	Analyzing how managers use intution and analytical tools to navigate complex problems and uncertainty	Making decisions based on analytics	Strategizing	B
7	Highest Prob: get, just, know, want, say, look, one, day, person, keep FREX: hous, tell, get, facebook, feel, doesnt, want, someth, keep, say	Monitoring personal behaviors to define digital marketing strategies customized to consumers	Monitoring customers' preferences	Understanding Customers	C
18	Highest Prob: custom, product, servic, consum, retail, brand, offer, store, segment, buy FREX: assort, loyalti, journey, flourish, merchandis, platform, buy, custom, seller, store	Understanding how customers' emotional connection with a brand relates to their behavior and how it impacts the organization	Monitoring customers' emotions	Understanding Customers	C
6	Highest Prob: cost, rate, growth, increas, valu, demand, per, exhibit, averag, expect FREX: per, margin, par, cost, increment, dollar, fix, variabl, exhibit, percentag	Using data analysis and statistics to predict the behavior of regional economies in the global scenario	Predicting global trends in economy	Market Forecasting	D
9	Highest Prob: product, price, qualiti, advertis, manufactur, sale, chang, relat, line, produc FREX: telemarket, qualiti, price, manufactur, defect, japanes, advertis, inventori, distributor, volum	Using data to evaluate the relationship between investments in advertisement and the sensitivity of price in determined markets	Indentifying the optimum price	Market Forecasting	D
12	Highest Prob: system, manag, plan, base, oper, order, process, resourc, control, use FREX: apex, oil, system, abc, iri, order, termin, hase, suppli, personnel	Using computational tools to perform project management related activities (i.e. budgeting, monitoring costs, stages of the project)	Managing processes with technology	Managing Processes	E

Topics (i.e., clusters of words from STM)

Axial Coding

Rodrigo Valadao, University of Alberta, August 7th 2020

Map of Correlations

Topics (from previous slide)

2

Year of publication

Aggregate Level	A Managing People	B Strategizing	C Understanding Customers	C Understanding Customers	B Strategizing	A Managing People	A Managing People	B Strategizing	B Strategizing	D Market Forecasting	D Market Forecasting	E Managing Processes
	Topic 1	Topic 3	Topic 18	Topic 7	Topic 11	Topic 19	Topic 14	Topic 5	Topic 15	Topic 6	Topic 9	Topic 12
2nd Order / Year	Promoting workplace diversity	Forstering research and innovation	Monitoring customers' emotions	Monitoring customers' preferences	Innovation shaping decisions	Evaluating team performance	Managing people based on analytics	Allocating resources	Making decisions based on analytics	Predicting global trends in economy	Indentifying the optimum price	Managing processes with technology
2018	0.06936 ***	0.07944 ***	0.07337 ***	0.07918 ***	0.03551 ***	0.04498 ***	0.01942 *	0.00282	0.01670 *	-0.05112 ***	-0.14532 ***	-0.22783 ***
2017	0.07653 ***	0.06398 ***	0.02706 *	0.04210 ***	0.04642 ***	0.01841 *	0.03575 ***	0.02398 **	0.01223	-0.04062 ***	-0.14568 ***	-0.22653 ***
2016	0.16777 ***	0.11799 ***	0.03071 **	0.04789 ***	0.02542 **	0.03873 ***	0.03283 ***	0.00605	0.00743	-0.05068 ***	-0.14694 ***	-0.23379 ***
2015	0.04016 ***	0.04344 ***	0.07583 ***	0.04321 ***	0.01987 **	0.01431 *	0.06237 ***	0.00235	-0.00442	-0.04328 ***	-0.14436 ***	-0.23069 ***
2014	0.03228 ***	0.06270 ***	0.08315 ***	0.04428 ***	0.04063 ***	0.01759 **	0.01340	0.02713 ***	0.03895 ***	-0.03964 ***	-0.13253 ***	-0.22168 ***
2013	0.04423 ***	0.07603 ***	0.04442 ***	0.05338 ***	0.03763 ***	0.04798 ***	0.02350 **	0.01463 *	0.02333 ***	-0.04544 ***	-0.13916 ***	-0.22099 ***
2012	0.04096 ***	0.04521 ***	0.06790 ***	0.04004 ***	0.02960 ***	0.01719 **	0.04711 ***	0.00844	0.00724	-0.01605	-0.13038 ***	-0.22630 ***
2011	0.04143 ***	0.02370 **	0.09366 ***	0.02787 **	0.02692 ***	0.00077	0.01401	0.02572 **	0.00054	0.00311	-0.13775 ***	-0.22686 ***
2010	0.09982 ***	0.06019 ***	0.03074 **	0.03094 ***	0.03474 ***	0.02366 **	0.02620 **	0.01304	0.02248 **	-0.03010 **	-0.13874 ***	-0.21963 ***
2009	0.10652 ***	0.03740 ***	0.03324 ***	0.03278 ***	0.04294 ***	0.02638 ***	0.01728 *	0.03627 ***	0.01107	-0.02480 *	-0.13738 ***	-0.22650 ***
2008	0.08586 ***	0.03216 ***	0.05257 ***	0.03107 ***	0.04196 ***	0.03404 ***	0.01353	0.02690 ***	0.01738 *	-0.02528 *	-0.13254 ***	-0.20958 ***
2007	0.05371 ***	0.03835 ***	0.05589 ***	0.07460 ***	0.04261 ***	0.04133 ***	0.02048 *	0.02547 ***	0.01008	-0.04576 ***	-0.13624 ***	-0.22061 ***
2005	0.01667	0.03897 ***	0.07641 ***	0.04222 ***	0.03844 ***	0.03384 **	0.01891	0.00119	0.00582	-0.04542 **	-0.13270 ***	-0.20283 ***
2004	0.03816 **	0.02406 **	0.03808 **	0.01941 *	0.01372	0.04408 ***	0.07565 ***	0.00532	-0.00757	-0.00116	-0.12612 ***	-0.22019 ***
2003	0.03770 ***	0.04646 ***	0.02260 *	0.04004 ***	0.02565 **	0.03315 ***	0.01234	0.01009	0.03211 ***	-0.03971 ***	-0.12626 ***	-0.21879 ***
2002	0.01335	0.13877 ***	0.01145	0.01460	-0.00239	0.09866 ***	0.01109	-0.00460	0.03286 **	-0.03726 *	-0.06205 **	-0.20528 ***
2001	0.02889 *	0.02648 **	0.04867 ***	0.02391 **	0.07691 ***	0.04166 ***	0.01270	0.03258 ***	0.01553	-0.03601 **	-0.12688 ***	-0.20541 ***
1998	0.01331	0.01354	0.03838 ***	0.01814 *	0.02006 **	0.01710 *	0.01728 *	0.00245	-0.00343	0.01110	-0.11339 ***	-0.12848 ***
1997	0.00796	0.01772	0.04673	0.03812	0.01195	0.00412	0.01092	0.00534	0.00610	-0.02178	-0.05101	-0.20490 ***
1996	0.01823	0.02304 **	0.05718 ***	0.02502 **	0.03414 ***	0.01350	0.00628	0.02197 **	0.00340	-0.01779	-0.09205 ***	-0.18910 ***
1995	0.01225	0.01674	0.01499	0.01273	0.02092 *	0.00788	0.02431 *	0.02339 *	0.01154	0.02467	-0.09431 ***	-0.18782 ***
1994	-0.00273	0.00681	0.01185	-0.00579	-0.01088	-0.01025	0.00320	-0.01019	0.00416	0.19148 ***	-0.06439	-0.15487 ***
1991	0.00961	0.01676	0.01088	0.01924	0.00343	0.01955 *	0.00192	0.00371	-0.00047	-0.02958 *	-0.08543 ***	-0.18612 ***
1990	0.01171	0.01112	0.01581	0.00488	0.00928	0.00189	0.03754 ***	0.00789	0.00197	-0.00471	-0.08961 ***	-0.15432 ***
1989	0.00247	0.00728	0.02605 *	0.00190	0.00636	-0.00390	0.02364 *	0.01681 *	-0.00206	0.00268	-0.03446 *	-0.12227 ***
1988	0.00737	0.03273 ***	0.00077	0.00199	-0.00085	0.00246	0.01437	0.00714	0.01452	0.01591	-0.07693 ***	-0.15322 ***
1986	0.00171	0.01657	0.02930 *	0.00996	0.01279	-0.00231	0.00396	-0.00504	0.01980	-0.00386	-0.02580	-0.14064 ***
1985	0.01218	0.00574	-0.00286	0.00808	-0.00928	0.00583	-0.00547	-0.00674	-0.01565	-0.04133 **	-0.06833 ***	-0.20902 ***
1984	0.01690	0.00905	-0.00338	0.03000 **	0.00076	0.01567 *	0.00967	0.03621 ***	-0.01223	0.04826 ***	-0.02429	-0.17937 ***
1983	0.00394	0.00139	-0.00241	0.00343	-0.01128	-0.00296	0.02256 *	0.00167	-0.01133	0.04130 **	0.09147 ***	-0.16589 ***
1982	0.00812	0.01543	-0.00261	0.03313 **	0.00808	0.02155 *	-0.00180	0.00604	0.01234	-0.02838 *	-0.09108 ***	-0.09772 ***
1981	0.00055	0.00484	-0.00505	-0.00144	0.00092	-0.00327	-0.00148	0.01013	0.01475	0.06577 ***	-0.05919 ***	-0.06594 ***
1980	0.00581	0.00139	-0.00813	-0.00536	-0.00184	0.00080	0.00135	0.05195 ***	-0.00948	0.14386 ***	-0.02796	-0.16891 ***
1979	-0.00170	-0.00042	-0.00206	-0.00698	-0.00364	-0.01067	-0.00406	0.00285	-0.00934	0.04460 ***	0.03090 *	-0.04393 **
(intercept)	0.00684	0.01013	0.01737 *	0.01862 **	0.03736 ***	0.02624 ***	0.02312 ***	0.03644 ***	0.04203 ***	0.06896 ***	0.15401 ***	0.24282 ***

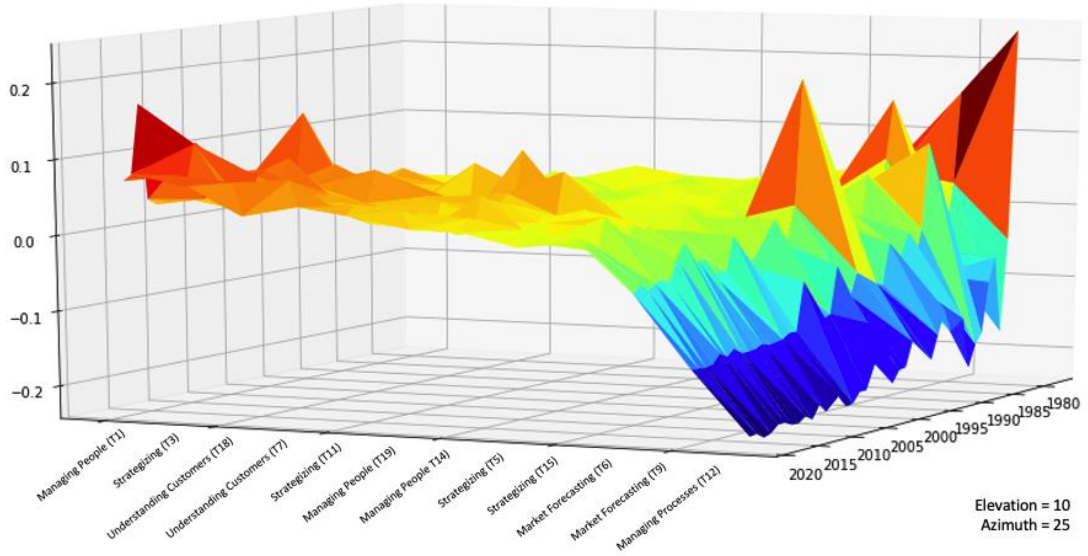
*p < .05; **p < .01; ***p < .001

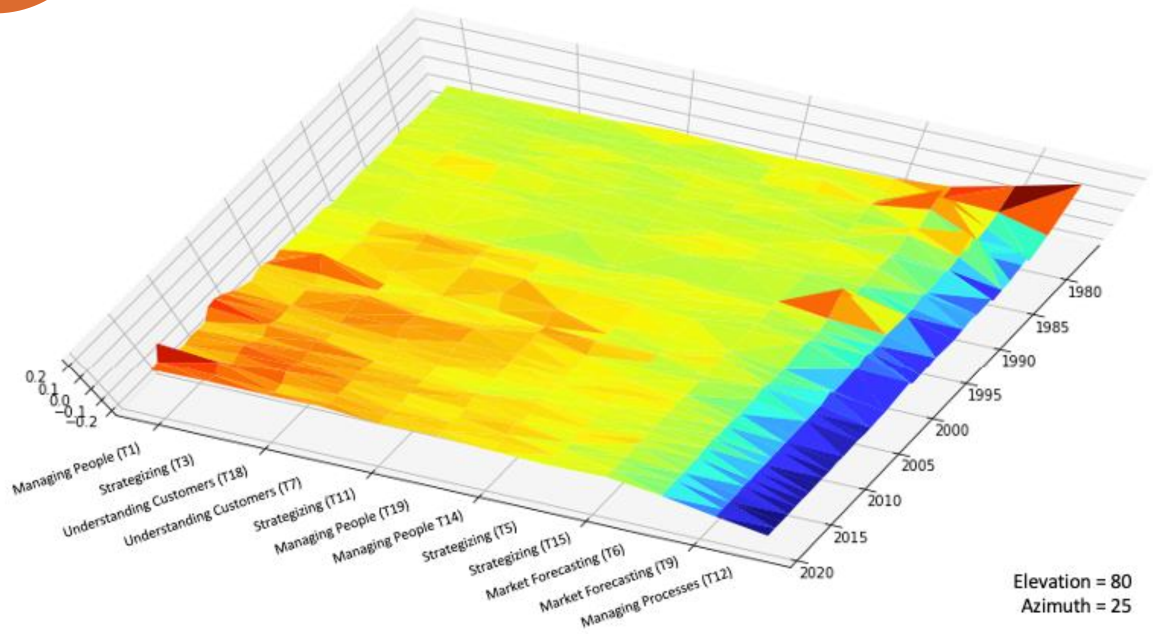
Greater than the mean for the topic

Smaller than the mean for the topic

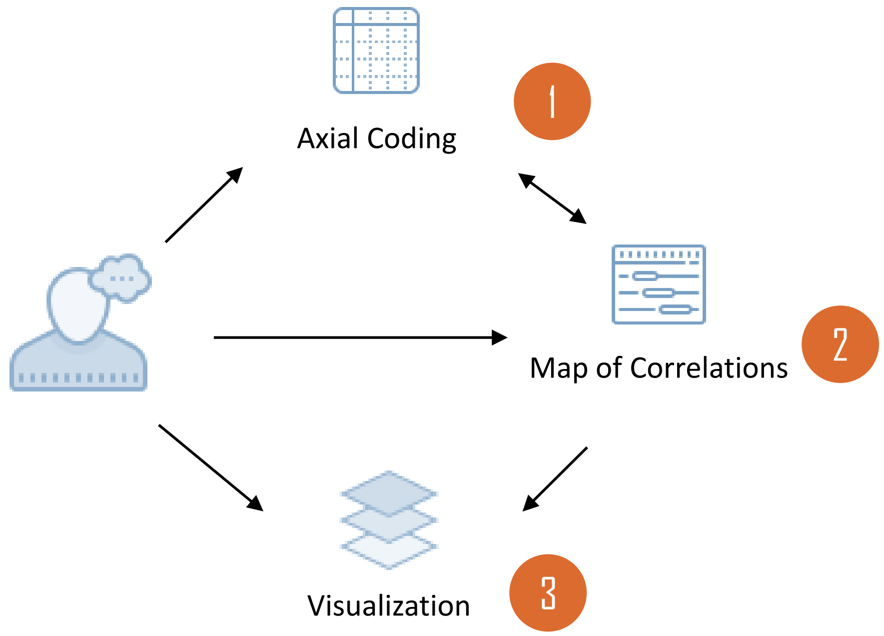
Visualization

3





Break-Out Session 1



Instructions

A. Each group will receive a handout with instructions and a copy of the three theoretical artifacts presented here.

B. **Key Task**: Discuss with your colleagues some potential interpretations of the theoretical artifacts and unfolding theorizations related to the research question: *How to capture the meaning infrastructure that shapes the space of possibilities at the early moments of institutional change?*

C. Pay special attention to patterns that might be identifiable into the artifacts , and take into consideration the following additional information:

In the interviews, most informants described 2010 as the year of emergence of Data Science as a field

D. After the break-out session, we will briefly present a potential theorization and hear back from the groups.

Rodrigo Valadao, University of Alberta, August 7th 2020

Discussion

- **Positive** and **negative** associations of elementary category schemas of meaning over time.
- 2000 x 2010: Divergence in the year detected by different approaches as the beginning of the emergence of the field (computational vs interviews).

Aggregate Level	A	B	C	C	B	A	A	B	B	D	D	E
	Managing People	Strategizing	Understanding Customers	Understanding Customers	Strategizing	Managing People	Managing People	Strategizing	Strategizing	Market Forecasting	Market Forecasting	Managing Processes
	Topic 1	Topic 3	Topic 18	Topic 7	Topic 11	Topic 19	Topic 14	Topic 5	Topic 15	Topic 6	Topic 9	Topic 12
Year \ 2nd Order	Promoting workplace diversity	Forstering research and innovation	Monitoring customers' emotions	Monitoring customers' preferences	Innovation shaping decisions	Evaluating team performance	Managing people based on analytics	Allocating resources	Making decisions based on analytics	Predicting global trends in economy	Indentifying the optimum price	Managing processes with technology
					+						-	
2018	0.06936***	0.07944***	0.07337***	0.07918	0.03551***	0.04498***	0.01942*	0.00282	0.01670*	-0.05112***	-0.14532	-0.22783***
2017	0.07653***	0.06398***	0.02706*	0.04210***	0.04642***	0.01841*	0.03575***	0.02398**	0.01223	-0.04062***	-0.14568***	-0.22653***
2016	0.16777***	0.11799***	0.03071**	0.04789***	0.02542**	0.03873***	0.03283***	0.00605	0.00743	-0.05068***	-0.14694***	-0.23379***
2015	0.04016***	0.04344***	0.07583***	0.04321***	0.01987*	0.01431*	0.06237***	0.00235	-0.00442	-0.04328***	-0.14436***	-0.23069***
2014	0.03228***	0.06270***	0.08315***	0.04428***	0.04063***	0.01759**	0.01340	0.02713***	0.03895***	-0.03964***	-0.13253***	-0.22168***
2013	0.04423***	0.07603***	0.04442***	0.05338***	0.03763***	0.04798***	0.02350**	0.01463*	0.02333***	-0.04544***	-0.13916***	-0.22099***
2012	0.04096***	0.04521***	0.06790***	0.04004***	0.02960***	0.01719**	0.04711***	0.00844	0.00724	-0.01605	-0.13038***	-0.22630***
2011	0.04143***	0.02370**	0.09366***	0.02787**	0.02692***	0.00077	0.01401	0.02572**	0.00054	0.00311	-0.13775***	-0.22686***
2010	0.09982***	0.06019***	0.03074**	0.03094***	0.03474***	0.02366**	0.02620**	0.01304	0.02248**	-0.03010**	-0.13874***	-0.21963***
2009	0.10652***	0.03740***	0.03324***	0.03278***	0.04294***	0.02638***	0.01728*	0.03627***	0.01107	-0.02480*	-0.13738***	-0.22650***
2008	0.08586***	0.03216***	0.05257***	0.03107***	0.04196***	0.03404***	0.01353	0.02690***	0.01738*	-0.02528*	-0.13254***	-0.20958***
2007	0.05371***	0.03835***	0.05589***	0.07460***	0.04261***	0.04133***	0.02048*	0.02547***	0.01008	-0.04576***	-0.13624***	-0.22061***
2005	0.01667	0.03897***	0.07641***	0.04222***	0.03844***	0.03384**	0.01891	0.00119	0.00582	-0.04542**	-0.13270***	-0.20283***
2004	0.03816**	0.02406**	0.03808**	0.01941*	0.01372	0.04408***	0.07565***	0.00532	-0.00757	-0.00116	-0.12612***	-0.22019***
2003	0.03770***	0.04646***	0.02260*	0.04004***	0.02565**	0.03315***	0.01234	0.01009	0.03211***	-0.03971***	-0.12626***	-0.21879***
2002	0.01335	0.13877***	0.01145	0.01460	-0.00239	0.09866***	0.01109	-0.00460	0.03286**	-0.03726*	-0.06205**	-0.20528***
2001	0.02889*	0.02648**	0.04867***	0.02391**	0.07691***	0.04166***	0.01270	0.03258***	0.01553	-0.03601**	-0.12688***	-0.20541***
1998	0.01331	0.01354	0.03838***	0.01814*	0.02006**	0.01710*	0.01728*	0.00245	-0.00343	0.01110	-0.11339***	-0.12848***
1997	0.00796	0.01772	0.04673	0.03812	0.01195	0.00412	0.01092	0.00534	0.00610	-0.02178	-0.05101	-0.20490***
1996	0.01823	0.02304**	0.05718***	0.02502**	0.03414***	0.01350	0.00628	0.02197**	0.00340	-0.01779	-0.09205***	-0.18910***
1995	0.01225	0.01674	0.01499	0.01273	0.02092*	0.00788	0.02431*	0.02339*	0.01154	0.02467	-0.09431***	-0.18782***
1994	-0.00273	0.00681	0.01185	-0.00579	-0.01088	-0.01025	0.00320	-0.01019	0.00416	0.19148***	-0.06439	-0.15487***
1991	0.00961	0.01676	0.01088	0.01924	0.00343	0.01955*	0.00192	0.00371	-0.00047	-0.02958*	-0.08543***	-0.18612***
1990	0.01171	0.01112	0.01581	0.00488	0.00928	0.00189	0.03754***	0.00789	0.00197	-0.00471	-0.08961***	-0.15432***
1989	0.00247	0.00728	0.02605*	0.00190	0.00636	-0.00390	0.02364*	0.01681*	-0.00206	0.00268	-0.03446*	-0.12227***
1988	0.00737	0.03273***	0.00077	0.00199	-0.00085	0.00246	0.01437	0.00714	0.01452	0.01591	-0.07693***	-0.15322***
1986	0.00171	0.01657	0.02930*	0.00996	0.01279	-0.00231	0.00396	-0.00504	0.01980	-0.00386	-0.02580	-0.14064***
1985	0.01218	0.00574	-0.00286	0.00808	-0.00928	0.00583	-0.00547	-0.00674	-0.01565	-0.04133**	-0.06833***	-0.20902***
1984	0.01690	0.00905	-0.00338	0.03000**	0.00076	0.01567*	0.00967	0.03621***	-0.01223	0.04826***	-0.02429	-0.17937***
1983	0.00394	0.00139	-0.00241	0.00343	-0.01128	-0.00296	0.02256*	0.00167	-0.01133	0.04130**	0.09147***	-0.16589***
1982	0.00812	0.01543	-0.00261	0.03313**	0.00808	0.02155*	-0.00180	0.00604	0.01234	-0.02838*	-0.09108***	-0.09772***
1981	0.00055	0.00484	-0.00505	-0.00144	0.00092	-0.00327	-0.00148	0.01013	0.01475	0.06577***	-0.05919***	-0.06594***
1980	0.00581	0.00139	-0.00813	-0.00536	-0.00184	0.00080	0.00135	0.05195***	-0.00948	0.14386***	-0.02796	-0.16891***
1979	-0.00170	-0.00042	-0.00206	-0.00698	-0.00364	-0.01067	-0.00406	0.00285	-0.00934	0.04460***	0.03090*	-0.04393**
(intercept)	0.00684	0.01013	0.01737*	0.01862**	0.03736***	0.02624***	0.02312***	0.03644***	0.04203***	0.06896***	0.15401***	0.24282***

Field Emergence

*p < .05; **p < .01; ***p < .001

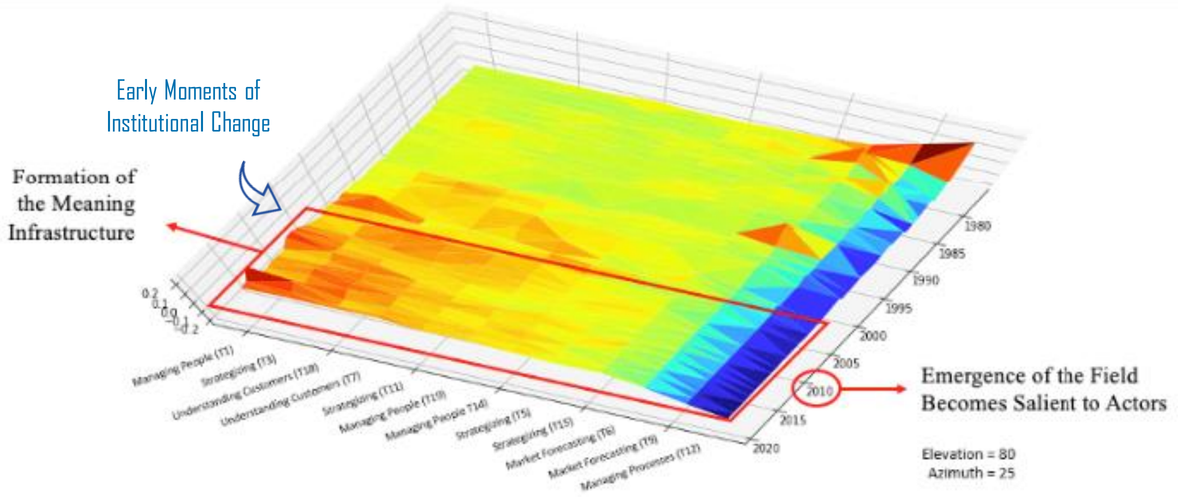
Greater than the mean for the topic

Smaller than the mean for the topic

Discussion

1

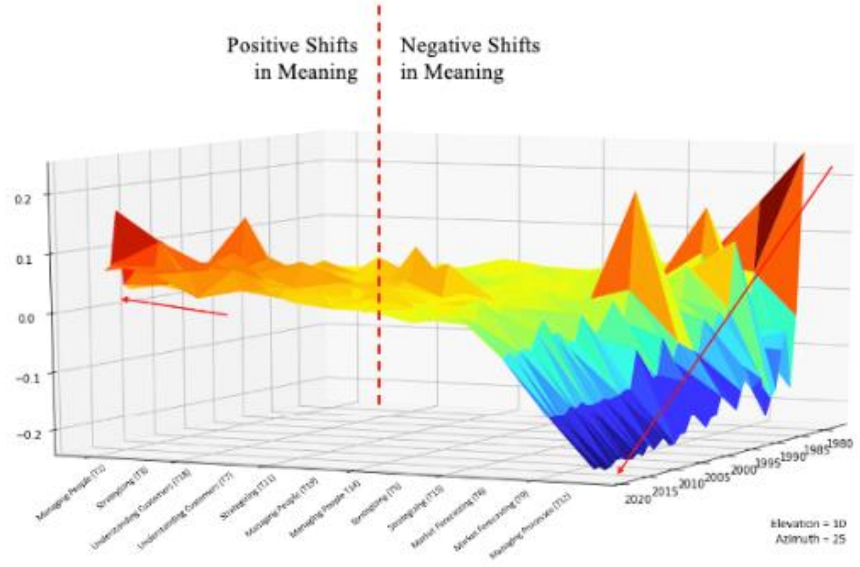
A meaning infrastructure is formed **through the amalgamation of novel category schemas of meaning in combination with the exclusion of others**



2

Distinct periods during the early moments of institutional change:

- **First period** (i.e., 2000-2010) → **tacit** and **not evident.**
- **Second Stage** (i.e., 2010-ongoing): might assign **a more agentic phase**, in which new practices start to become more closely connected to the meaning infrastructure.



The Big Picture about STM

- Meaning is a key component to understand change but has been difficult to operationalize it empirically (Mohr et al., 2020).

STM enables to develop new types of visualizations and opens an array of possibilities for novel theorizations.

Future Directions:

- Studies can **increase the variety of metadata** used as covariates in the STM technique. **Authorship**, for example, shall enable to populate studies of meaning with a more agentic and multidimensional perspective .

- Future studies might want to **take into account the pace of change.** The angular coefficient of the correlations produced by the STM technique might afford this sort of analysis, which requires further development of the technique.

Moving Beyond LDA & Standard Topic Modeling – hSBM

Example / Exercise Part 2

(Hovig Tchalian & Tim Hannigan)

ACADEMY OF Management

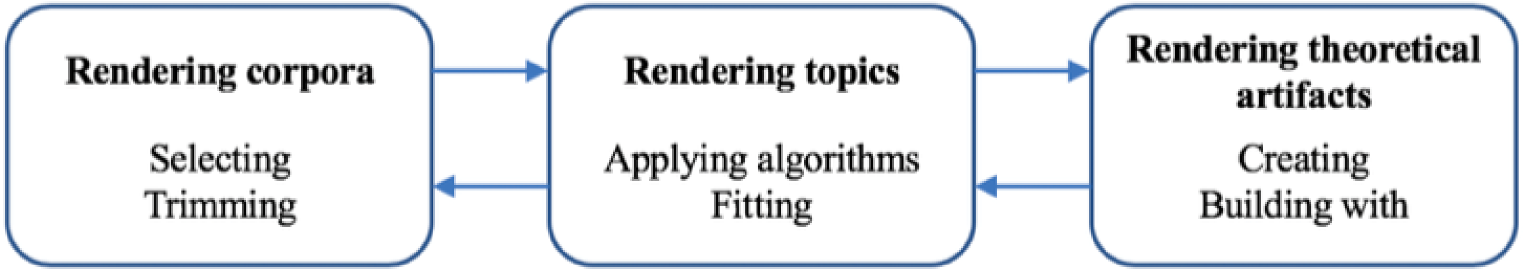
#AOM2020

80th Annual Meeting of the Academy of Management

Hannigan, Haans, Glaser, Tchalian, Valadao, Jennings
IDeaS
August 7, 2020

Building Up Visuals for Theorization
Case Application & Exercise

Figure 3
Topic Modeling Rendering in Theory-Building Spaces



Can also **identify** temporal **breaks**

Must be **rigorous** + **interpretable**

Visuals help build **theory**

Hannigan et al. (2019). Topic modeling in management research: Rendering new theory from textual data. *Academy of Management Annals*, 13(2), 586-632,

hLDA As Visual Artifact: *Malaysia Flight 370*

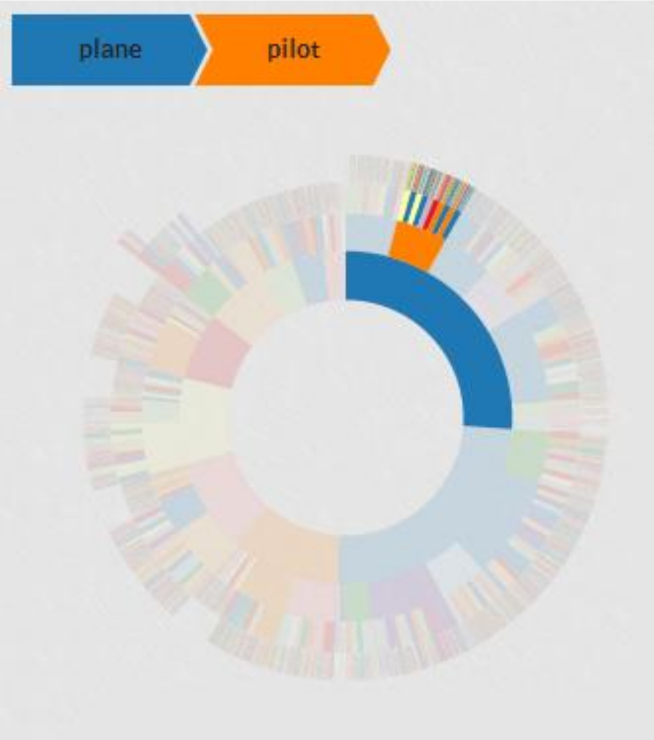


Figure 3: Our simple breadcrumb trail and contextual anchor offer constant context as the user explores the visualization. Highlighted slices within the contextual anchor are those currently displayed in the sunburst visualization.

plane, crash, crashed
plane, landed, land
plane, think, people
pilot, plane, hijacking
terrorist, terrorism, passports
suicide, pilot, ocean
Shah, Anwar, political
plane, China, world
phone, phones, cell
evidence, think, make

Table 1: The 10 high-level topics of the model generated from running HLDA on the Malaysia Flight MH-370 corpus. The bolded topics suggest specific theories regarding the status of the plane.

crash, water, crashed
failure, catastrophic, mayday
mechanical, failure, days
plane, ocean, did
plane, error, lost

Two Alternative (and Distinct) *Hierarchical* Models

hLDA

Supervised / top-down model

- Researcher chooses # of levels

  (but not # of topics)

- Topics can be difficult to label / interpret

- Visualizations can be crude

  (excel + sunburst diagram)

Blei, Griffiths & Jordan, The nested Chinese restaurant process and Bayesian nonparametric inference of topic hierarchies (Journal of the ACM, Vol. 57, No. 2, Article 7 2010)

hSBM

Unsupervised, network-based model

- Model chooses number of topics and number of levels (optimizes both parameters)

- Topics more interpretable + traceable (docs < > terms)

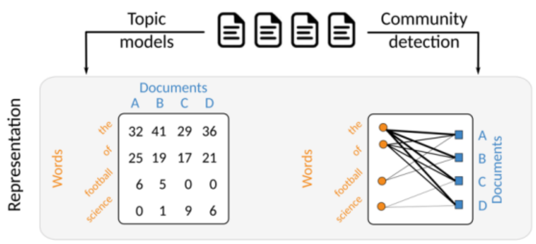
- Visualization more dynamic (esp. with code-level access)

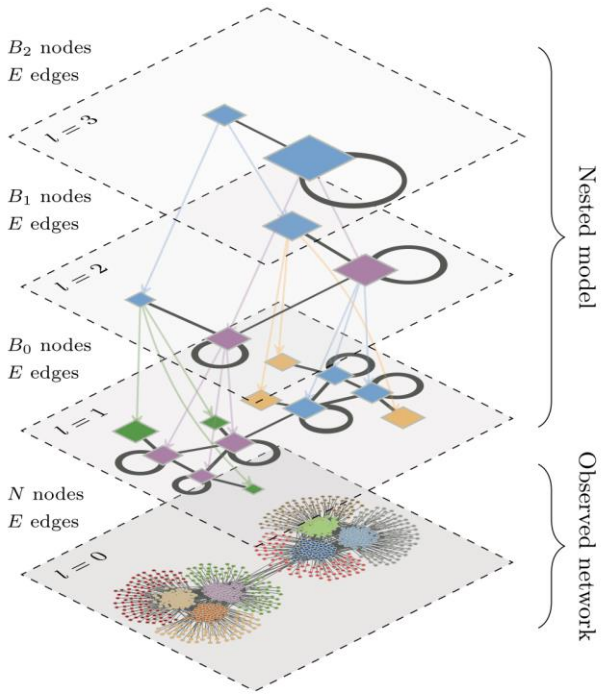
- Overcomes limitations of LDA models with natural language; ie, burstiness, correlations between topics

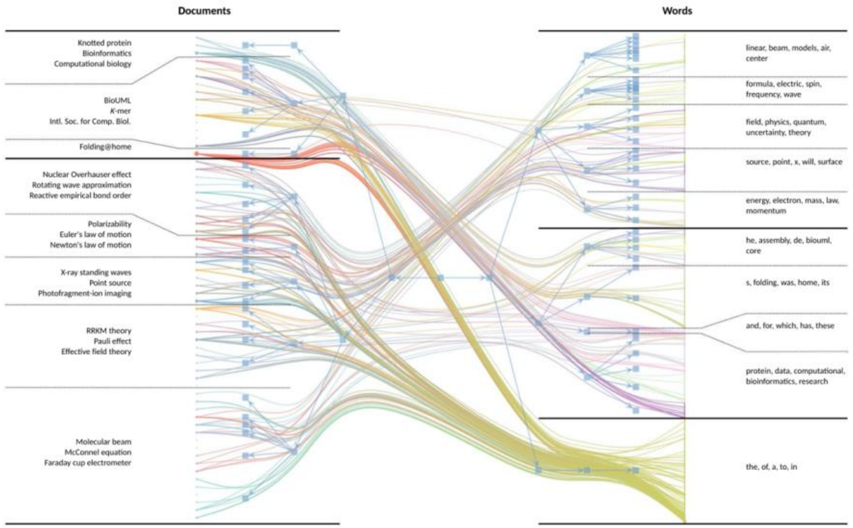
HSBM Offers Powerful Visualization + Theorization

A network approach to topic models -> hierarchical stochastic block modeling (hSBM)

- Canonical papers:
  - Gerlach, M., Peixoto, T. P., & Altmann, E. G. (2018). A network approach to topic models. *Science Advances*, *4*(7), eaaq1360.
  - Peixoto, T. P. (2014). Hierarchical Block Structures and High-Resolution Model Selection in Large Networks. *Physical Review X*, *4*(1), 011047.
- hSBM software (code): https://github.com/martingerlach/hSBM_Topicmodel/blob/master/TopSBM-tutorial.ipynb

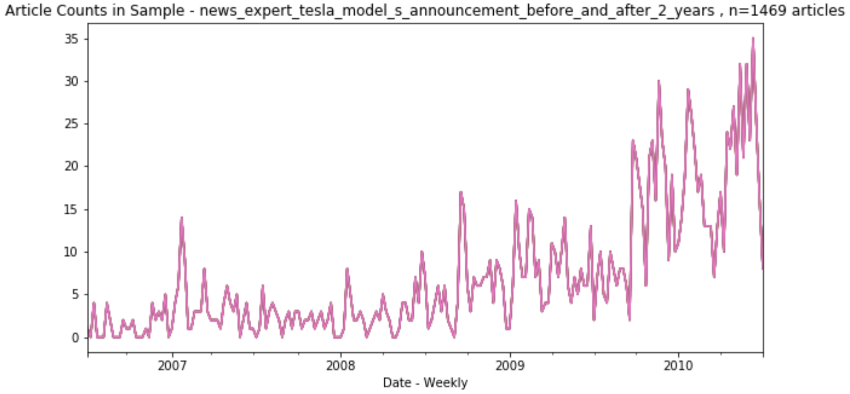


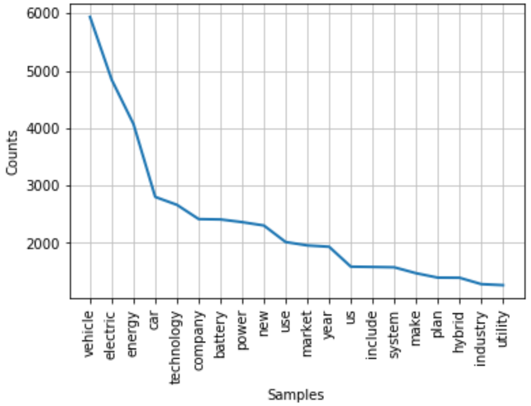




Case: Emergence of the Electric Vehicle Market

- Exploratory Research Question: *Is there an inflection point in the EV market emergence where the discourse changes?*
- Following the Rosa et al. (1999) study on the minivan market, we track discourse from expert publications in the Electric Vehicle market
- First exploration: whether the announcement of the Tesla Model S (June 30, 2008) corresponded with a discourse change



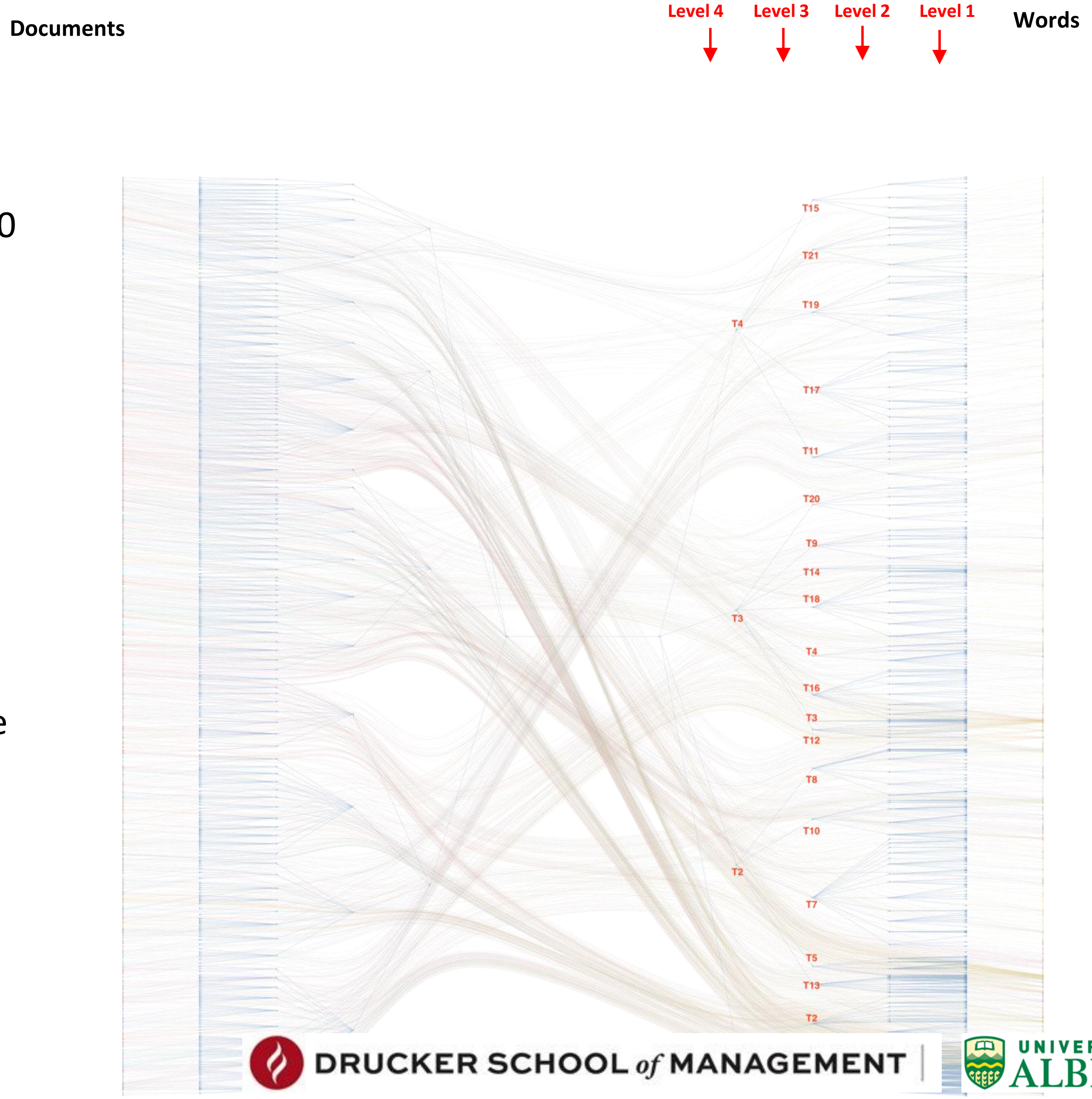


- Simple frequency analysis of article counts is promising, showing increase over time; but content analysis is inconclusive

DRUCKER SCHOOL *of* MANAGEMENT | UNIVERSITY OF ALBERTA

Case: Emergence of the Electric Vehicle Market

- We used two years of textual data (articles from expert publications) before and after the Model S announcement (2006-06-30 -> 2010-06-30)
- The figure on the right shows the hSBM visual output
- There are 4 levels of topics in this hierarchy
- We manually read the topics at the top level of abstraction, then went down to level 3 which were meaningful
- We then tracked topic salience over time (next slide)



Break-Out Session 2

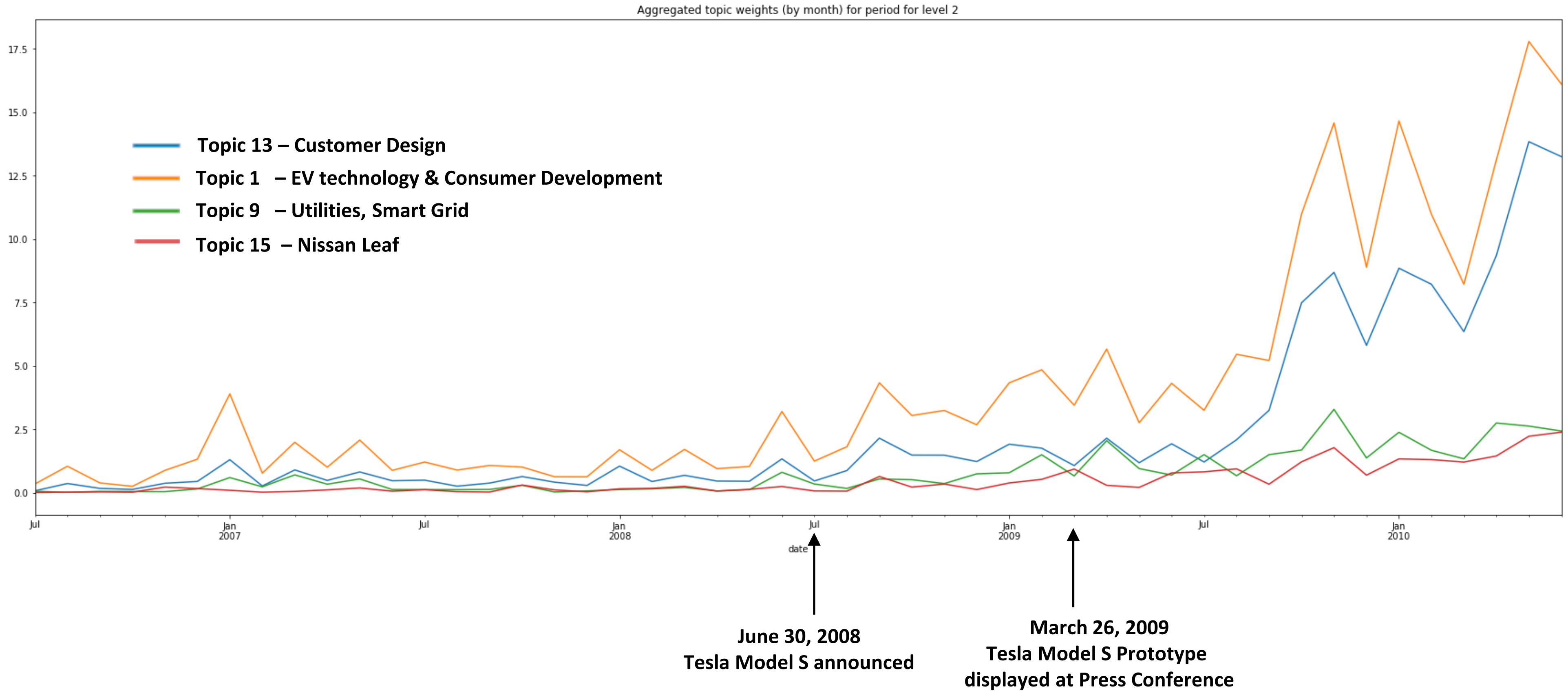
- Excel includes three artifacts:

  1. hSBM visualization
  2. hSBM topics outputs
     - Top 2 levels: 4 + 3
     - Topic-specific terms
     - Ranked by weight (high >> low)
  3. Sheets for your additions

- **Follow Excel instructions:**

  1. Use hSBM visual to inspect topics
  2. Label L4 topics + L3 sub-topics, iterating back to L4
  3. Identify preliminary theoretical insights

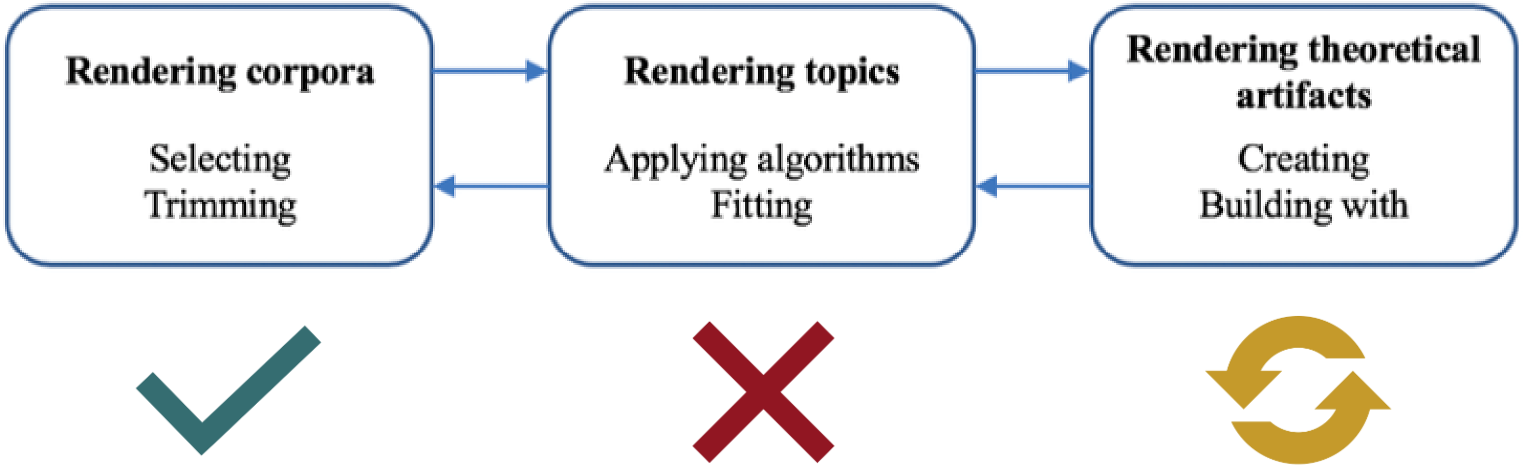
Discussion: Emergence of the EV Market



- Because overall article counts increased in the latter part of the period, we would expect to see some overall increase in topic salience over time
- From this figure, we see that Customer Design, EV Technology and Consumer Development topics increased after the Model S announcement
- Topics around Utilities, Smart Grid and the Nissan Leaf (EV) increased far slower

Theorization Is a Dynamic, Iterative Process

Figure 3
Topic Modeling Rendering in Theory-Building Spaces



Hannigan et al. (2019). Topic modeling in management research: Rendering new theory from textual data. *Academy of Management Annals*, 13(2), 586-632,

Rendering with STM & hSBM Enhances Visuals

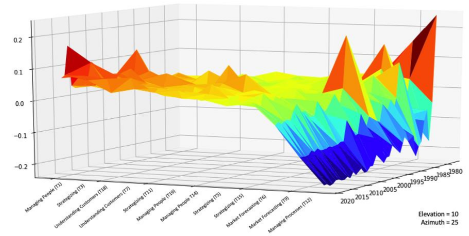
- Theories of institutions, culture, relationality and neo-structuralism rely on visuals and can be enhanced by rendering of visuals with topic modeling.
  - As artifacts (especially symbols) in cultures (likely enhanced in terms of capturing centrality of key cultural artifacts)
  - As boundary objects in field relations (likely, because discourse strands become more evident)
  - As representations of deeper structure (in combo with LDAviz, become more powerful)
  - As rhetorical devices (very likely, adding to particular storylines or types of rhetoric…)
  - As improved measures of extant concepts (already true, based on examples)

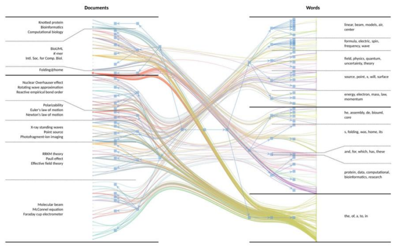
Topic Modeling Is An *Interpretive* Data Science (IDS)

- LDA and other implementations of topic modeling identify *latent structure*, based on a (dirichlet) probability distribution
- But generating insights requires a healthy dose of interpretation
- IDS combines quantitative *and* qualitative insights
- Methods more advanced than LDA allow for a more dynamic, iterative process of theorization
- Visualization can act as a critical aid to the theorization process

What's Next?

- IDeaS Workshop late 2020 or early 2021 = our "big tent" community
- New, and update: IDeaS general page: http://www.interpretivedatascience.com/
- Updated GitHub: https://ideas-repo.github.io
- Special Issue, we hope. ☺
- We hope that you folks don't mind **being signed up to our community**. If you don't want any info on the IDeaS event, please let me know.





Topic Modeling Advances

Thank You!

ACADEMY OF Management

#AOM2020

80th Annual Meeting of the Academy of Management

Hannigan, Haans, Glaser, Tchalian, Valadao, Jennings
IDeaS
August 7, 2020