ADVANCED CONTENT ANALYSIS TECHNIQUES

Content Analysis PDW
Academy of Management – 2014
Philadelphia, PA

Tim Hannigan (University of Oxford)
Robert Vesco (Yale University)

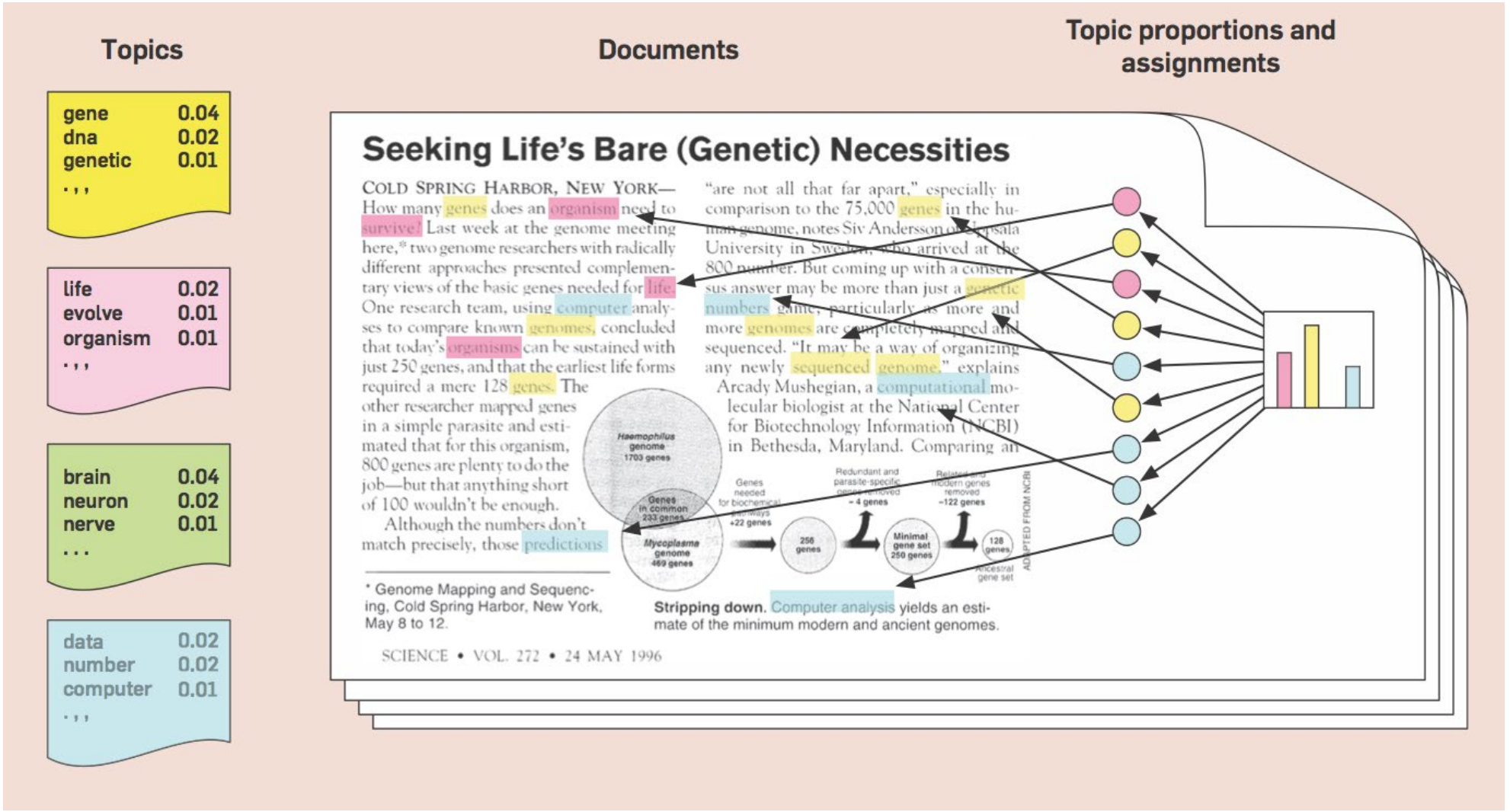
GOALS

- Share examples of cutting edge text analysis techniques by running through a sample case
- Discuss challenges and future promise
- Provide tips and resources for implementation

ADVANCED TECHNIQUES

- **Topic Models**
- Name Entity Recognition (NER)
- High Accuracy Sentiment Analysis
- Concept Networks

TOPIC MODELING OVERVIEW (BLEI, 2011)



TOPIC ANALYSIS

- What is it?
  - Generates probabilistic models of topic/categories within text
  - Most commonly Latent Dirichlet Allocation , or LDA
- When is it useful? Examples?
  - Looking at the change in ideas over time (Kaplan & Vakili, 2014; Mohr et al, 2013; DiMaggio et al, 2013)
  - Identifying relationships between entities based on shared meanings

Example: AOM PDWs Abstracts

- We web-scraped 2014 AOM PDWs
- Texts of Abstracts
- Sponsors (BPS, OMT, COG, TIM …)
- Extracted “topics” using LDA Topic Modeling
- Analysed relationship between topics and sponsors

Printed Program Preview

Program Session #: **32** | Submission: **11536** | Sponsor(s): **(MOC, RM, OMT, BPS, SIM, OB)**
Scheduled: **Friday, Aug 1 2014 8:00AM - 12:30PM** at **Loews Philadelphia Hotel** in **Commonwealth C**

View Map

Content Analysis in Organizational Research: Techniques and Applications
Content Analysis Research

Coordinator: **Moriah A. Meyskens**; U. of San Diego;
Coordinator: **Michael D. Pfarrer**; U. of Georgia;
Presenter: **Michael K. Bednar**; U. of Illinois;
Facilitator: **Jonathan Bundy**; Pennsylvania State U.;
Presenter: **Timothy R. Hannigan**; U. of Oxford;
Presenter: **Jason Kiley**; U. of Georgia;
Facilitator: **Aaron Francis McKenny**; U. of Central Florida;
Facilitator: **Vilmos F. Misangyi**; Pennsylvania State U.;
Presenter: **Todd W. Moss**; Syracuse U.;
Facilitator: **Rhonda K. Reger**; U. of Tennessee;
Facilitator: **Robert Vesco**; Robert H. Smith School of Business;
Facilitator: **Miles A. Zachary**; West Virginia U.;

This two-part PDW runs from 8-10 AM and 10:30-12:30 PM on Friday, August 1. Part 1 provides an introduction to content analysis as a research methodology. Presenters will discuss appropriate applications, reliability and validity concerns, and different computer-aided content analysis tools. Experts will also walk through examples of content analysis techniques from published research and offer publishing tips. Part 1 is open to all AOM attendees and does not require pre-registration. Part 2 of the PDW models MOC's successful "Cognition in the Rough" PDW. Experts and authors will interact in small groups to discuss the content, structure, techniques, and potential journal outlets of submitted proposals. Part 2 requires pre-registration and a submission of a proposal to contentanalysis1@gmail.com. The deadline for proposal submissions is June 15. Details are available via AOM listservs or by contacting the organizers at contentanalysis1@gmail.com. The Oxford Centre for Corporate Reputation will graciously sponsor the PDW, provide refreshments, and host a reception for all attendees.

Search Terms: Content and text analysis , Qualitative and quantitative , Method

Tweet this session: #AOM2014 32

KEY TO SYMBOLS
Teaching-oriented | Practice-oriented | International-oriented | Theme-oriented | Research-oriented | Diversity-oriented
Selected as a Best Paper

AOM PDWs

(AAA) All Academy Activities
(AAT) All Academy Theme
(AAC) Affiliate Activities & Committees
(AAM) Asia Academy of Management
(BPS) Business Policy & Strategy
(CAR) Careers
(CAU) Caucuses
(CM) Conflict Management
(CMS) Critical Management Studies
(D&ITC) Diversity & Inclusion Theme Committee
(ENT) Entrepreneurship
(EXH) Exhibits
(GDO) Gender & Diversity in Organizations
(HCM) Health Care Management
(HR) Human Resources
(IAM) Iberoamerican Academy of Management
(ICW) In Conjunction With Activities
(INDAM) Indian Academy of Management
(IM) International Management
(ITC) International Theme Committee
(MC) Management Consulting
(MED) Management Education & Development
(MH) Management History
(MSR) Management Spirituality & Religion
(MOC) Managerial & Organizational Cognition
(NDSC) New Doctoral Student Consortium
(OM) Operations Management
(OMT) Organization & Management Theory
(ODC) Organization Development & Change
(OB) Organizational Behavior
(OCIS) Organizational Communication & Information Systems
(ONE) Organizations & the Natural Environment
(PTC) Practice Theme Committee
(PNP) Public & Nonprofit
(RM) Research Methods
(SIM) Social Issues in Management
(SAP) Strategizing Activities and Practices
(TLC) Teaching & Learning Conference
(TTC) Teaching Theme Committee
(TIM) Technology & Innovation Management

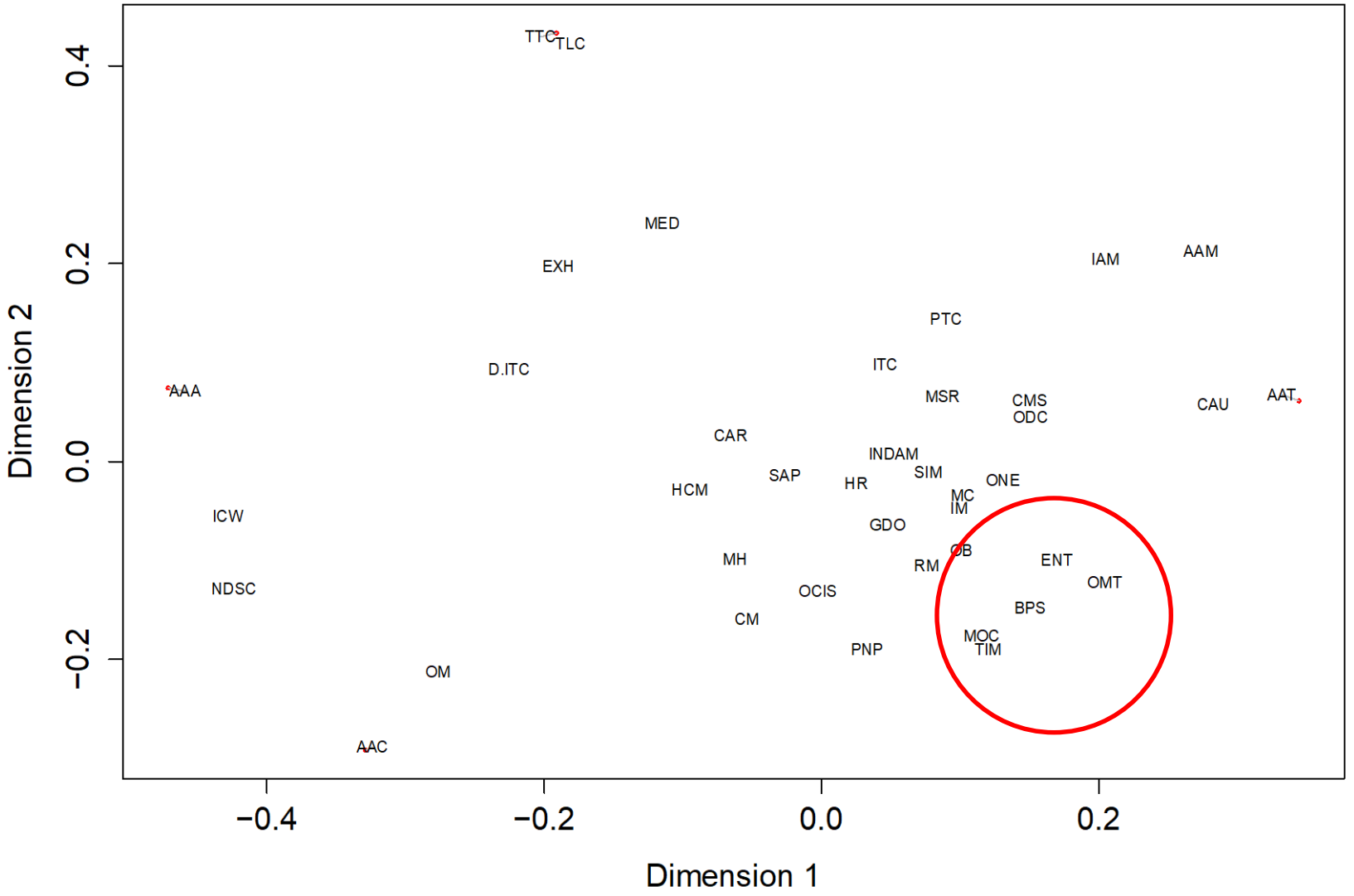
NEED TO INTERPRET TOPICS

Topic 11	Topic 15	Topic 20
"social"	"entrepreneuri”	"gender"
"ethic"	"ventur"	"women"
"csr"	"entrepreneursh	"career"
"stakehold	"entrepreneur"	"divers"
"respons"	"new"	"femal"
"corpor"	"busi"	"negoti"
"moral"	"opportun"	"studi"
"studi"	"startup"	"work"
"manag"	"studi"	"find"
"busi"	"find"	"organ"
"paper"	"model"	"effect"
"valu"	"effect"	"men"
"theori"	"research"	"research"
"practic"	"theori"	"differ"
"find"	"firm"	"manag"
"engag"	"uncertainti"	"posit"
"uneth"	"use"	"experi"
"use"	"paper"	"use"
"firm"	"capit"	"male"
"public"		"age"

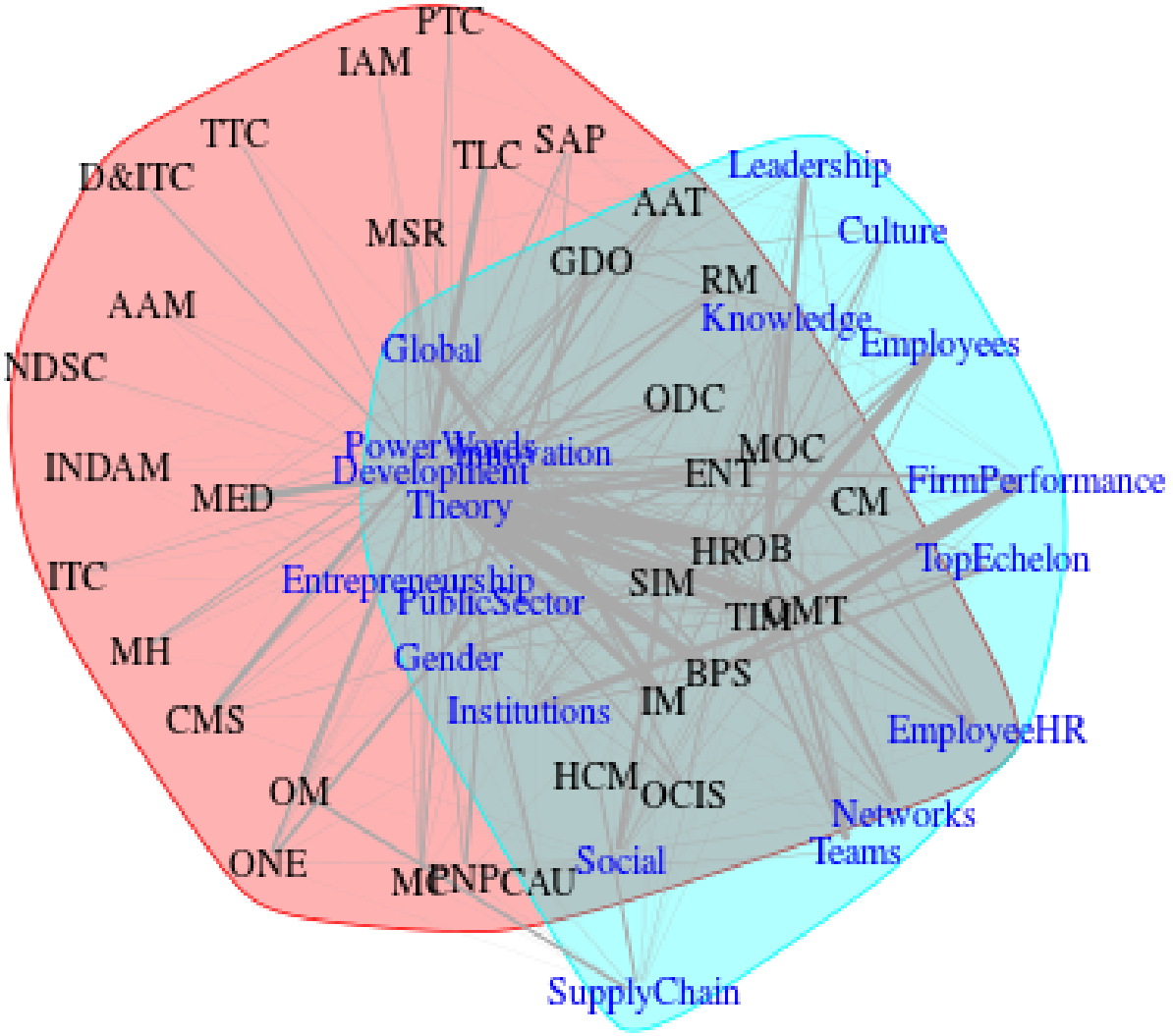
TOP TOPICS BY DIVISION

	BPS	OB	MOC	OMT	TIM
EmployeeHR	2.00	28.00	11.00	6.00	0.00
FirmPerformance	46.00	1.00	0.00	5.00	41.00
Innovation	9.00	5.00	1.00	9.00	18.00
Institutions	11.00	1.00	6.00	37.00	1.00
Theory	50.00	62.00	20.00	48.00	28.00

RELATEDNESS BY DIVISION



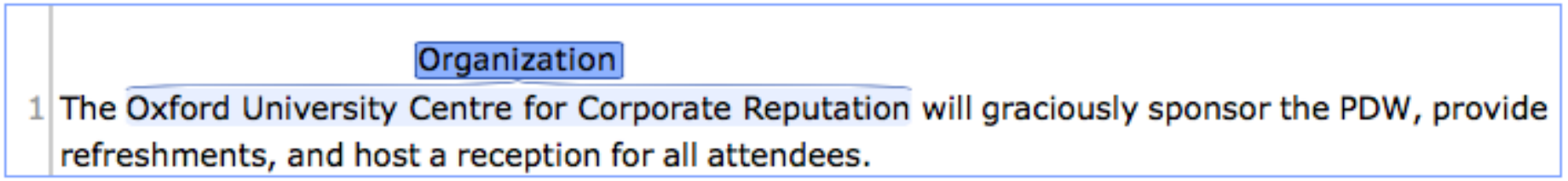
NETWORKS


PTC
IAM
TTC
TLC
SAP
D&ITC
Leadership
AAT
MSR
Culture
GDO
RM
AAM
Knowledge
Employees
Global
NDSC
ODC
PowerWords
Innovation
MOC
INDAM
Development
ENT
FirmPerformance
MED
Theory
CM
HROB
ITC
TopEchelon
Entrepreneurship
SIM
PublicSector
TIM
OMT
MH
Gender
BPS
IM
CMS
Institutions
EmployeeHR
HCM
OCIS
OM
Networks
Teams
ONE
Social
MC
NPCAU
SupplyChain

NAMED ENTITY RECOGNITION

What is it?

Find names (people and places) within text



When is it useful? Examples?

You want to identify individuals or geographic locations mentioned
Allows you to do social network analysis and spatial econometrics
Currently using to identify government agencies in patent data

Advanced Sentiment Analysis

- **What is it?**
  - Is the text negative, positive or neutral; can bring Natural Language Processing, so negation of valence can be captured.. perhaps even sarcasm (yeah, right!)
  - Machine learning

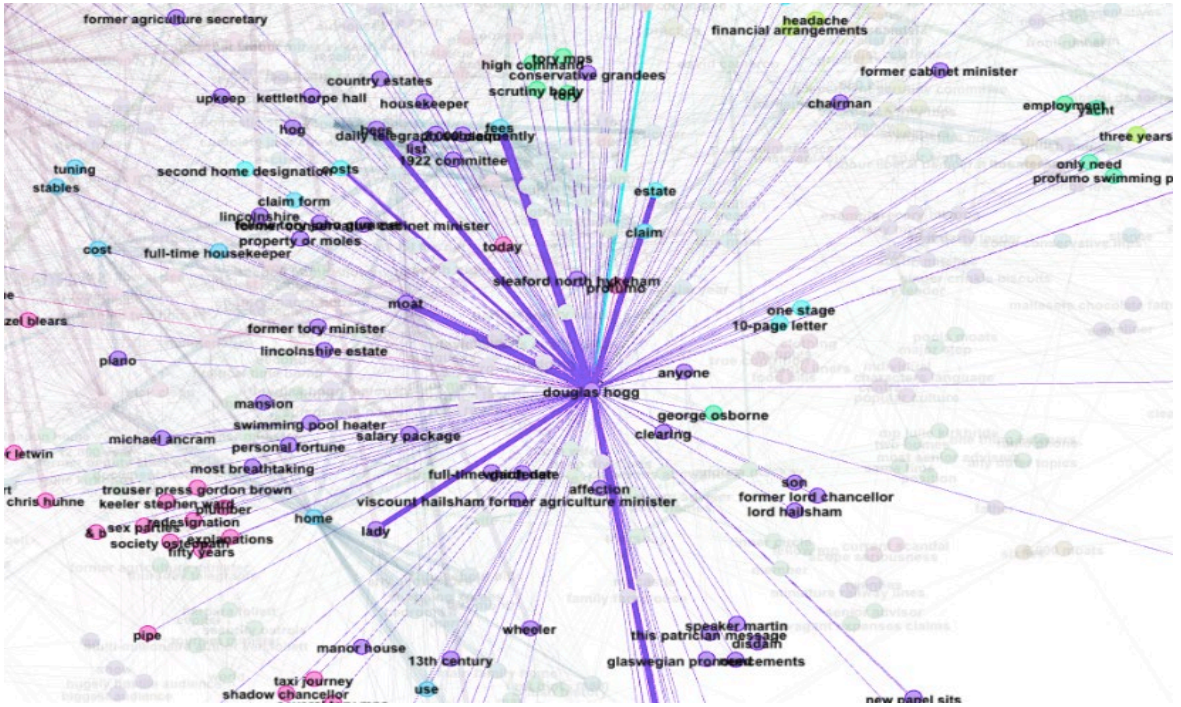
- **Advantages**
  - Same as always – but many high-end services provide high accuracy (>90%)
  - Going beyond word dictionaries and simple word counts, moving beyond LIWC (Kaplan, 2011)

- **Downsides:**
  - Methodological blackbox
  - Complexity in process / cost

- **Services will do it for you**
  - Have access to large corpora (ie. Google Books, NGrams), language is evolving (ie. Google, big data)
  - Relatively cheap … to not so much

Concept Networks

- Can use collocations of concepts to form a network; then can use network tools such as centrality to measure salience
- To be covered more in **PDW "Revisiting the Product Ontology"** (Sat Aug 2, 10:14-12:45, Pennsylvania Convention Center Room 203B) and **Symposium "The Power of Words in Big Data"** (Sun Aug 3 11:15-12:45, Pennsylvania Convention Center, Room 122 A)



NEXT CHALLENGES

- Promise
  - there's no out of the box tool to do this for you
  - there are opportunities to collaborate with computer science researchers
  - opportunities to integrate this with Network Analysis Tools
  - Ethnographers and computers scientists can work together using topic modelling and complement one another (eg. Levy & Franklin, 2013)
  - emergent properties (large amounts of data)
- Challenges:
  - tools can be a black box; may be sensitive to certain assumptions
  - despite their scale and speed, there still remains a lot of researcher degrees of freedom

RESOURCES

- Stanford Topic Modeling Toolbox - http://www-nlp.stanford.edu/software/tmt/
- Topic Modeling in R using LDA
- Topic Modeling Tutorial in R and Python
- http://java.dzone.com/articles/topic-modeling-python-and-r
- Python programming for the Humanities - http://fbkarsdorp.github.io/python-course/
- Text analysis with topic models for the Humanities and social sciences
- https://de.dariah.eu/tatom/index.html
- https://github.com/rlvesco7/aom2014-content-analysis-pdw