GET THE FACT(OID)S STRAIGHT

TOWARDS DEEPER VERISIMILITUDE ESTIMATION

KARO MOILANEN

TheySay Ltd

Department of Computer Science, University of Oxford

Sentiment Analysis Symposium | 15-JUL-2015

ACKNOWLEDGEMENTS

Dr Simon Guest

Prof Stephen Pulman

*"Is Greece like going bankcrupt or some shit?! Cameron n Merkel is probabli blah blah blah tevs facepalm
It'll be Germanys downfall..so wotEvA #DILLIGAS #euro Zzzzzz
5 more days until i travel to GREEZE yay"*

*"BBC: Greece faces a critical 24 hours as an emergency summit is held today.
Cameron, Merkel, and most other Eurozone leaders want to break the deadlock around the country's debt crisis."*

GRICE'S COOPERATIVE PRINCIPLE

A rough, general communication principle which participants in an effective rational conversation will be expected to observe

> *"Make your conversational contribution*
> *such as is required,*
> *at the stage at which it occurs,*
> *by the accepted purpose or*
> *direction of the talk exchange*
> *in which you are engaged."*

Grice, H. P. (1975): 45.

GRICE'S CONVERSATIONAL MAXIMS

Quantity	• Make your contribution as informative as is required (for the current purposes of the exchange). • Do not make your contribution more informative than is required.
Quality	• Try to make your contribution one that is true. • Do not say what you believe to be false. • Do not say that for which you lack adequate evidence.
Relation	• Be relevant.
Manner	• Be perspicuous. • Avoid obscurity of expression. • Avoid ambiguity. • Be brief (avoid unnecessary prolixity). • Be orderly.

Grice, H. P. (1975): 45-8.

VERISIMILITUDE

- Given that the maxims are flouted or violated, some texts are true(r) to the maxims (unmarkedness), some less so (markedness)
- We quantify qualitative markedness with a complex, subjective, and probabilistic composite measure along a continuum between

⊕	truthful | truthlike | fact	plausible | possible	clear	relevant
⊖	untruthful | untruthlike | factoid	implausible | impossible	unclear	irrelevant

TRUTHFULNESS

- Truthful: the speaker's intention not to misrepresent information
- Untruthful: deliberate or unintentional misrepresention of information by the speaker

	Speaker	Hearer
Truth	true	true
Truth	false	false
Lie	true	false
Lie	false	true
BS	*	*

TRUTHFULNESS

TOPOLOGIES OF DECEPTION

- Distortions: lies | exaggerations
- Unlies: false implications | misleadings
- Concealments: secrets | half-truths | masks
- Diversionary responses: hedges | nonsequiturs | evasion | topic switches | irrelevance | ambiguity | equivocation | amphiboly | vagueness | doublethink | accent
- Other: crimes | fictions | playings

(after Hopper & Bell (1984) and Turner et al. (1975) in Burgoon (1996))

PLAUSIBILITY

- Not all events or states of affairs in the world are equally
  - frequent | possible
- Not all thoughts, ideas, propositions, beliefs, and assumptions are equally
  - conventional | thinkable
- Not all logical deductions and inferences are equally
  - provable | inferrable | consistent
- Therefore, texts differ across these dimensions, too

CLARITY AND RELEVANCE

- Since not all texts require the same amount of cognitive processing, they vary in
  - understandability | readability | well-formedness | coherence | consistency | naturalness
- Discourse structure, presentation, and delivery all shape verisimilitude greatly
  - flow | cohesion | interestingness
- Depending on the topic, turn or stage in the discourse, or the speaker or hearer in question, texts vary tremendously in
  - topicality | aboutness | informativeness

LINGUISTIC VERISIMILITUDE COMPUTATION

- A complex composite measure comes with many big linguistic detection, classification, and scoring challenges
- Propositional truth | inference | presupposition
- Entailment | factivity
  - Lotan et al. (2013) | Androutsopoulos & Malakasiotis (2010) | MacCartney et al. (2006)
- Contradiction
  - de Marneffe et al. (2008) | Ritter et al. (2008)
- Veridicality | veridicity | uncertainty | beliefs | modality | speculation | hedging
  - de Marneffe et al. (2011) | Moncecchi et al. (2012) | Farkas et al. (2010) | Sanchez & Vogel (2015) | Szarvas et al. (2012) | Karttunen & Zaenen (2005)

LINGUISTIC VERISIMILITUDE COMPUTATION

- Non-literal meaning | metaphors | figurative language
  - Shutova (2015) | Loenneker-Rodman & Narayanan (2008)
- Sarcasm | irony
  - Bamman & Smith (2015) | Ghosh et al. (2015) | Barbieri & Saggion (2014) | Lukin & Walker (2013) | Reyes et al. (2013)
- Humour
  - Zhang & Liu (2014) | Kao et al. (2013) | Mihalcea & Pulman (2007)
- Understandability | readability | information quality
  - Collins-Thompson (2014) | Flor et al. (2013) | Pitler & Nenkova (2008) | Kate et al. (2010)

LINGUISTIC VERISIMILITUDE COMPUTATION

- Bias | framing
  - Baumer et al. (2015) | Recasens et al. (2013)
- Redundancy | text simplification
  - Horn et al. (2014) | Zanzotto et al. (2011)
- Objectivity | factuality | subjectivity
- Sentiment | emotion | affect

EXTRALINGUISTIC VERISIMILITUDE COMPUTATION

- Even more challenges stem from extralinguistic factors
- Information credibility | reliability

    Castillo et al. (2011) | Mitra & Gilbert (2015)
- Fact checking
    - Ciampaglia et al. (2015)
- User-specific relevance and interestingness criteria
- Information propagation | memes | rumours
    - Qazvinian et al. (2011)
- World knowledge | ultimately everything...

VERISIMILITUDE COMPUTATION: APPROACHES

- Predictably, the majority of studies surrounding the topic have resorted to (un)supervised learning
- Need for rich features for specific topics, cognitive dimensions, syntax, stylistics, and discourse
- OK performance in some but not all aspects of verisimilitude
- (Un)availability of training data
  - plenty of 'proper' vs. 'junk' quality texts available
  - topic-specific deception data sets typically come from artificial, simulated lab conditions
- Hence no panacea in challenging real-world conditions

VERISIMILITUDE COMPUTATION: APPROACHES

- Some verisimilitude cues can be modelled directly
  - deception cues identified in psychological studies
  - many simple but well-established readability scores and text clarity measures
  - while some are moderately robust, many cues are too weak (individually) or even contradictory
  - substractive ‘lie’ vs. additive ‘truth’ cues
  - easily configurable
- Some form of inductive reasoning is required
- Large knowledge graphs and bases
  - fact checking and plausibility estimates

HUMAN PERFORMANCE (ACCURACY?)

> *"If liars were much better,*
> *truth telling would be less common:*
> *if detectors were much better,*
> *few lies would be attempted."*

Bond & De Paulo (2006): 233.

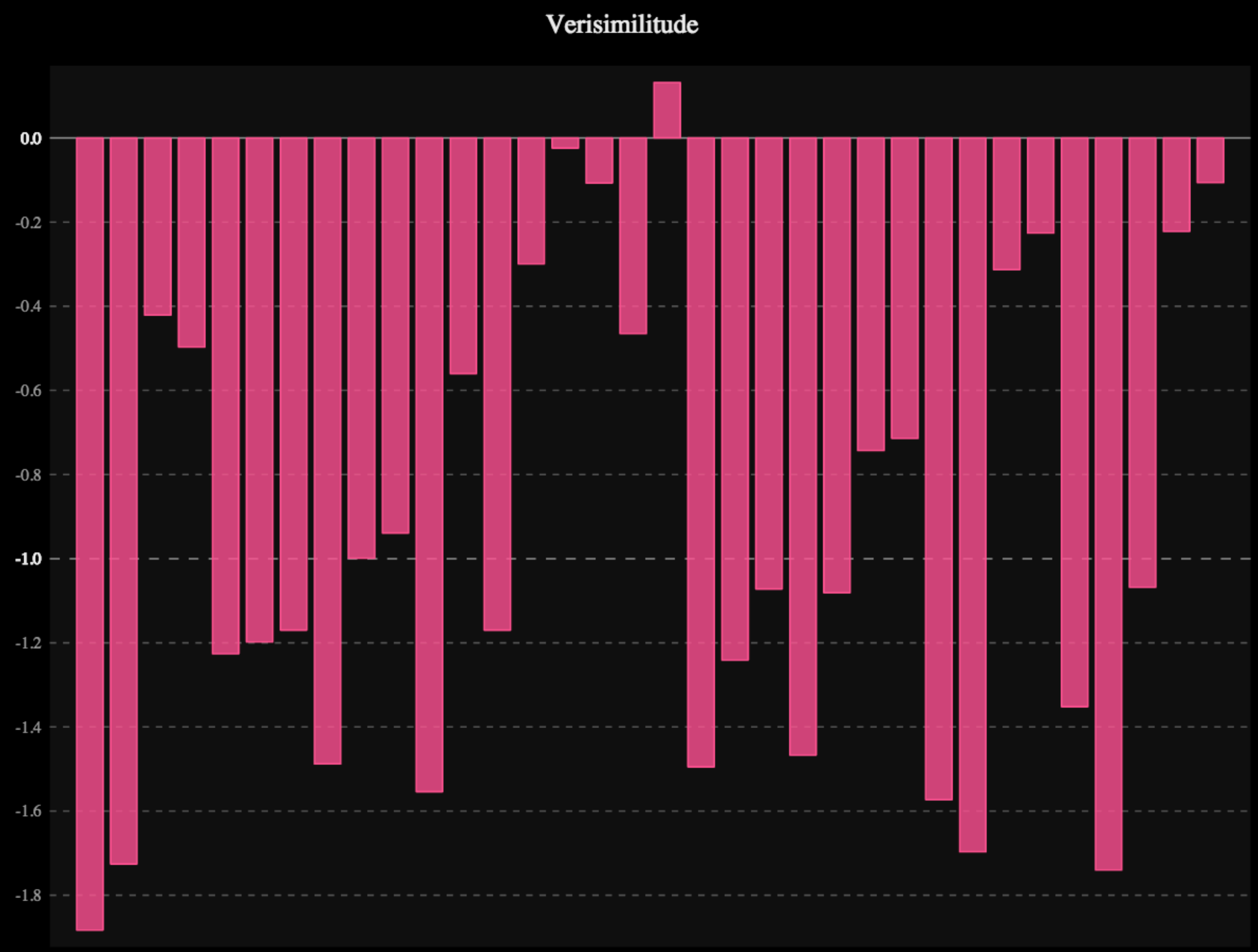
HUMAN PERFORMANCE (ACCURACY?)

34% acc	Recognising lies	Levine (2006)
40.7% acc	Spotting biased words	Recasens et al. (2013)
54% ~ 54.5% acc (meta-study)	Recognising deception	Aamodt (2006), Bond & DePaulo (2006)
57.5% ~ 65% acc	Recognising negative opinion spam	Ott (2013)
.63 *k*	Recognising reduntant tweets	Zanzotto et al. (2011)
67% acc	Recognising truth	Levine (2006)
.77 ICC	Scoring credibility	Mitra & Gilbert (2015)
.81 *k*	Recognising contradictions	Marneffe (2008)
.95 *k*	Extracting rumours	Qazvinian et al. (2011)
91~96% acc	Interpreting entailment	MacCartney et al. (2006)

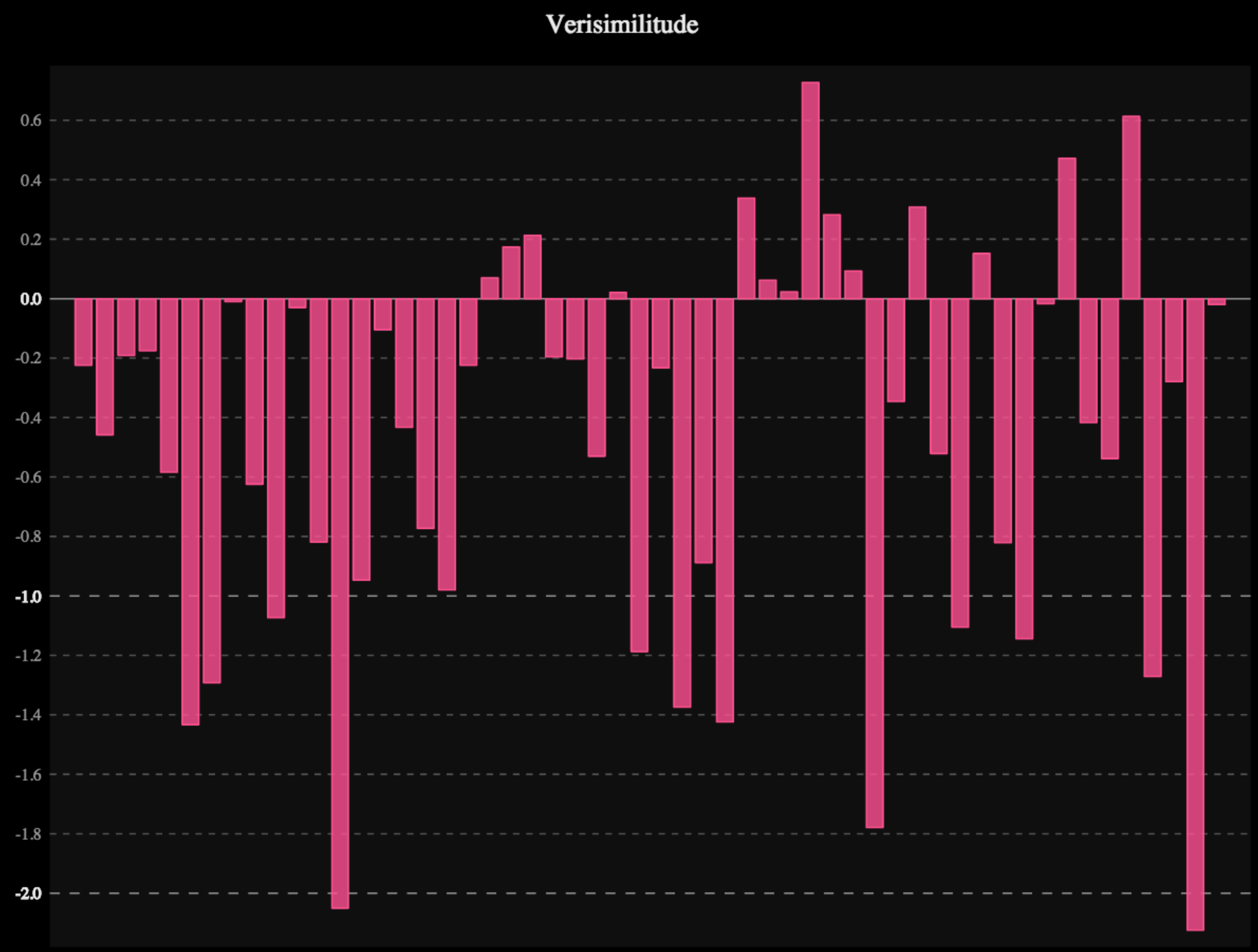
SAMPLE DATA

- UK 2015 General Election
- Four UK politicians
- Short speeches and manifestos by the politicians
    - Spring 2015
- 1062 tweets by the politicians
    - January - June 2015
- 1m tweets mentioning the politicians
    - random sample from 3 029 276 tweets from January till May 15 2015
    - 250k tweets per politician

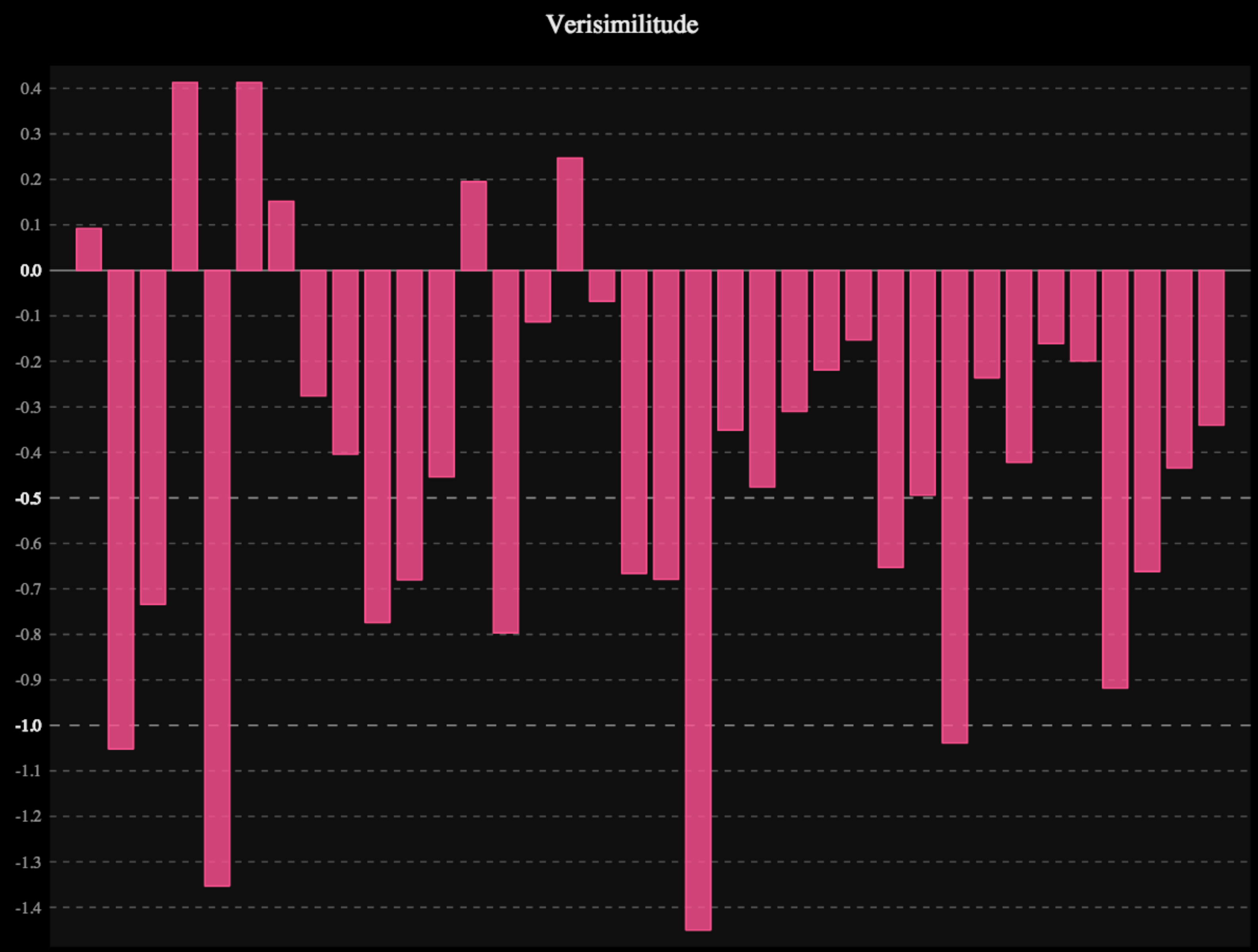
SPEECH BY POLITICIAN 1



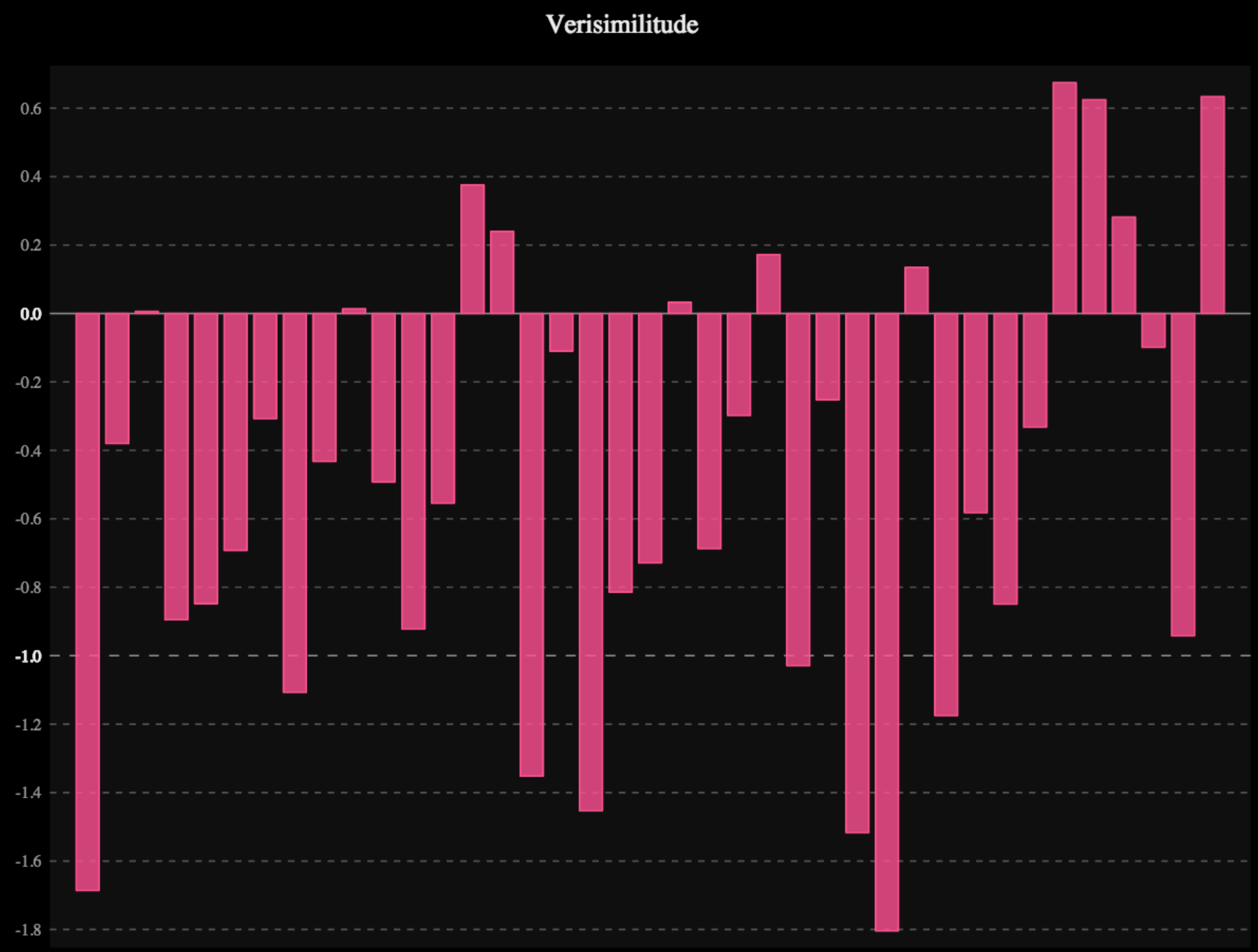
SPEECH BY POLITICIAN 2



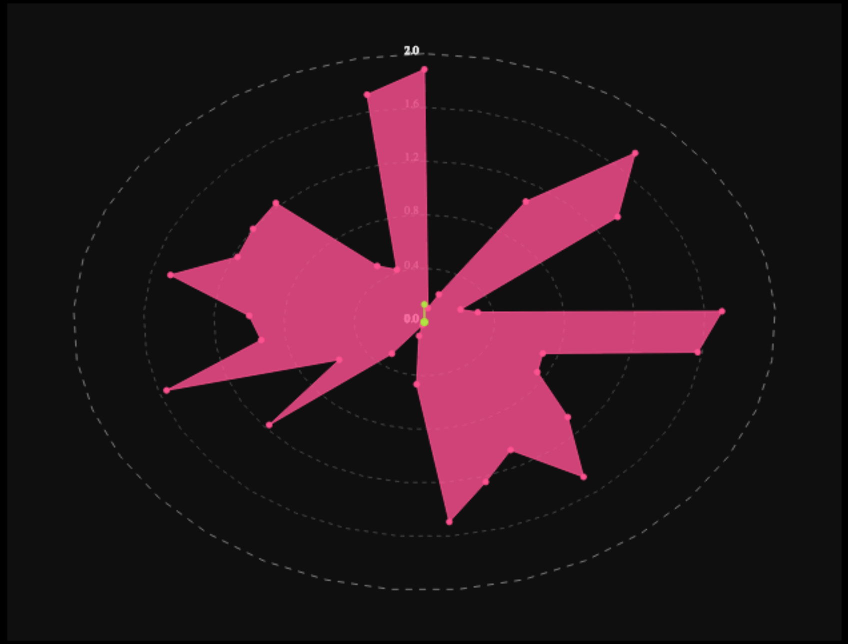
SPEECH BY POLITICIAN 3

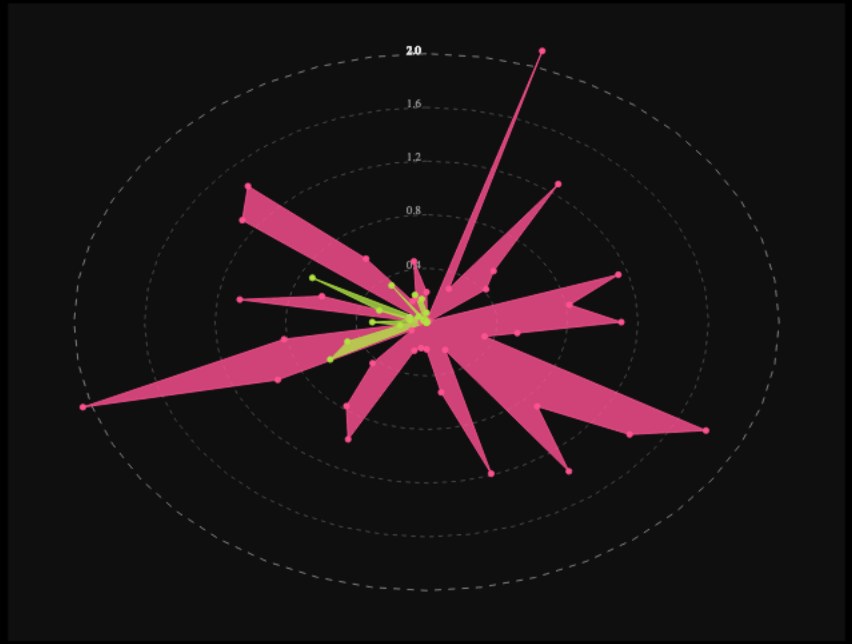


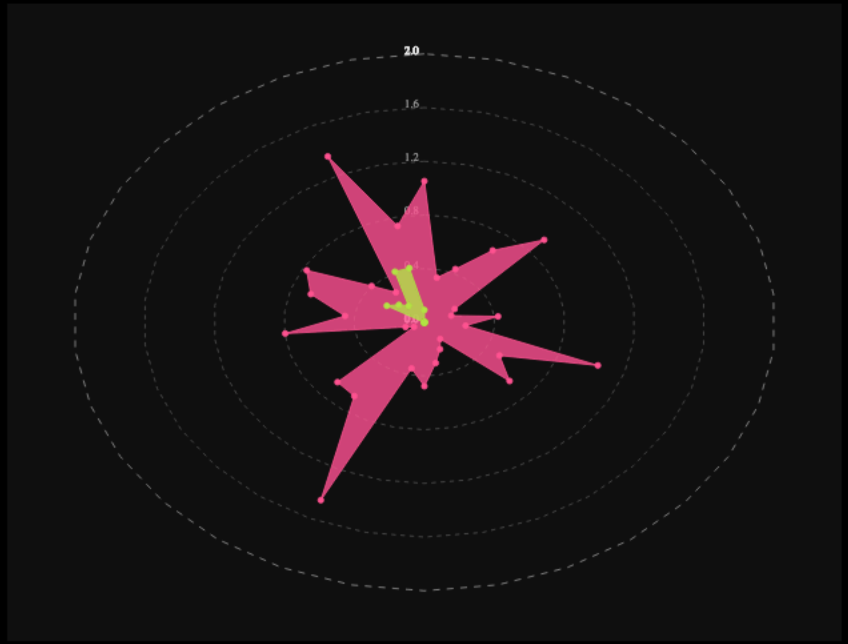
SPEECH BY POLITICIAN 4

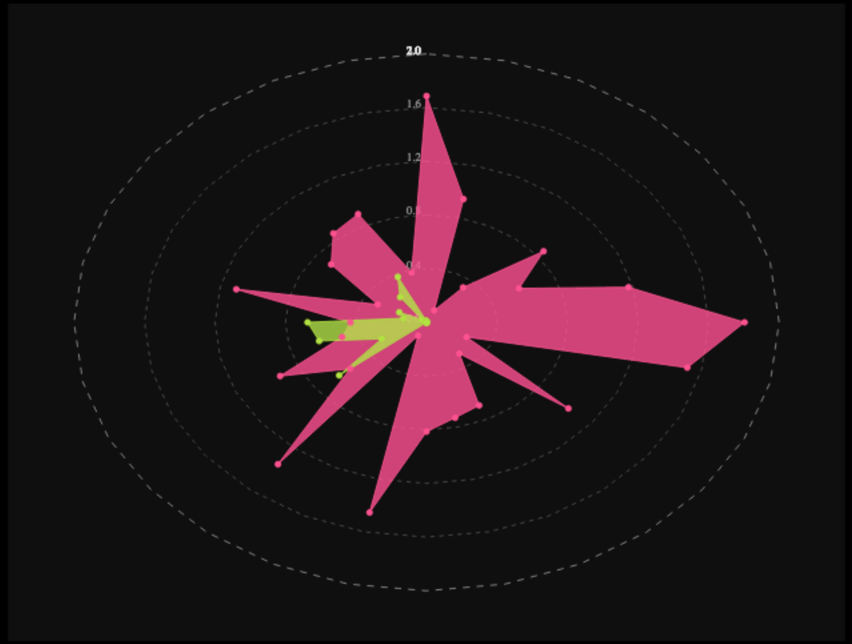


SPEECHES BY POLITICIANS 1-4



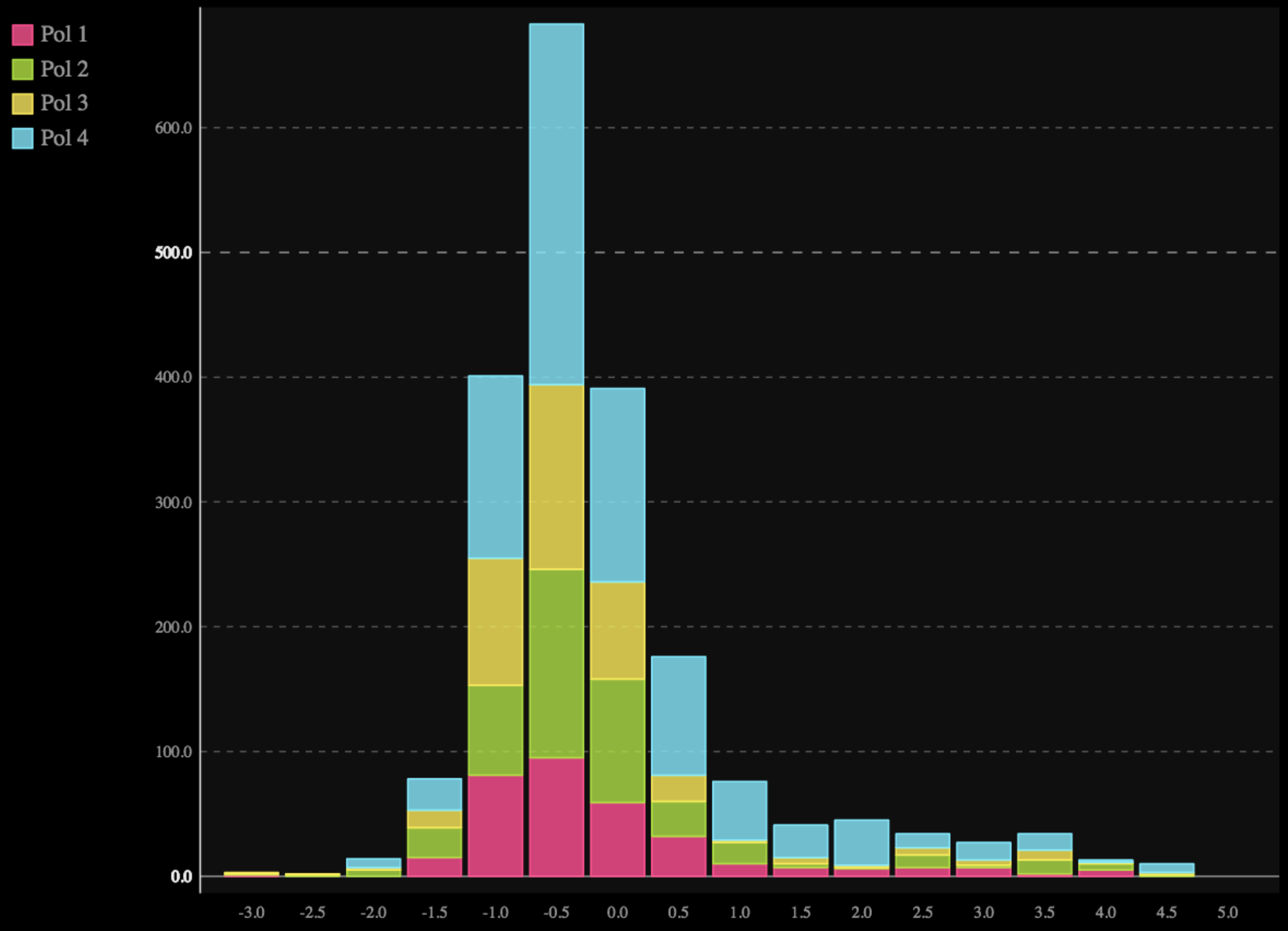






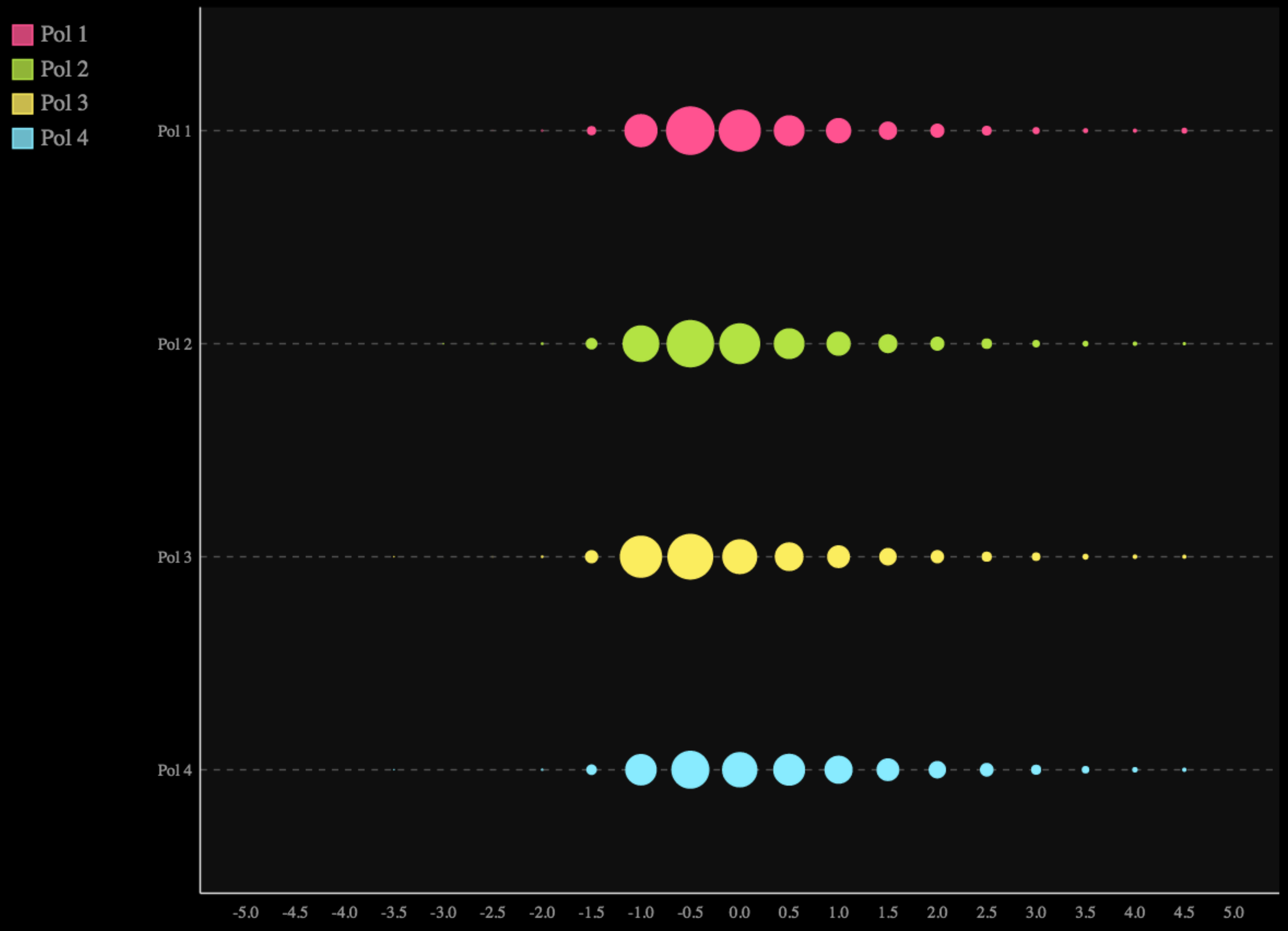
DISTRIBUTION OF SCORES

1k Tweets by Politicians 1-4



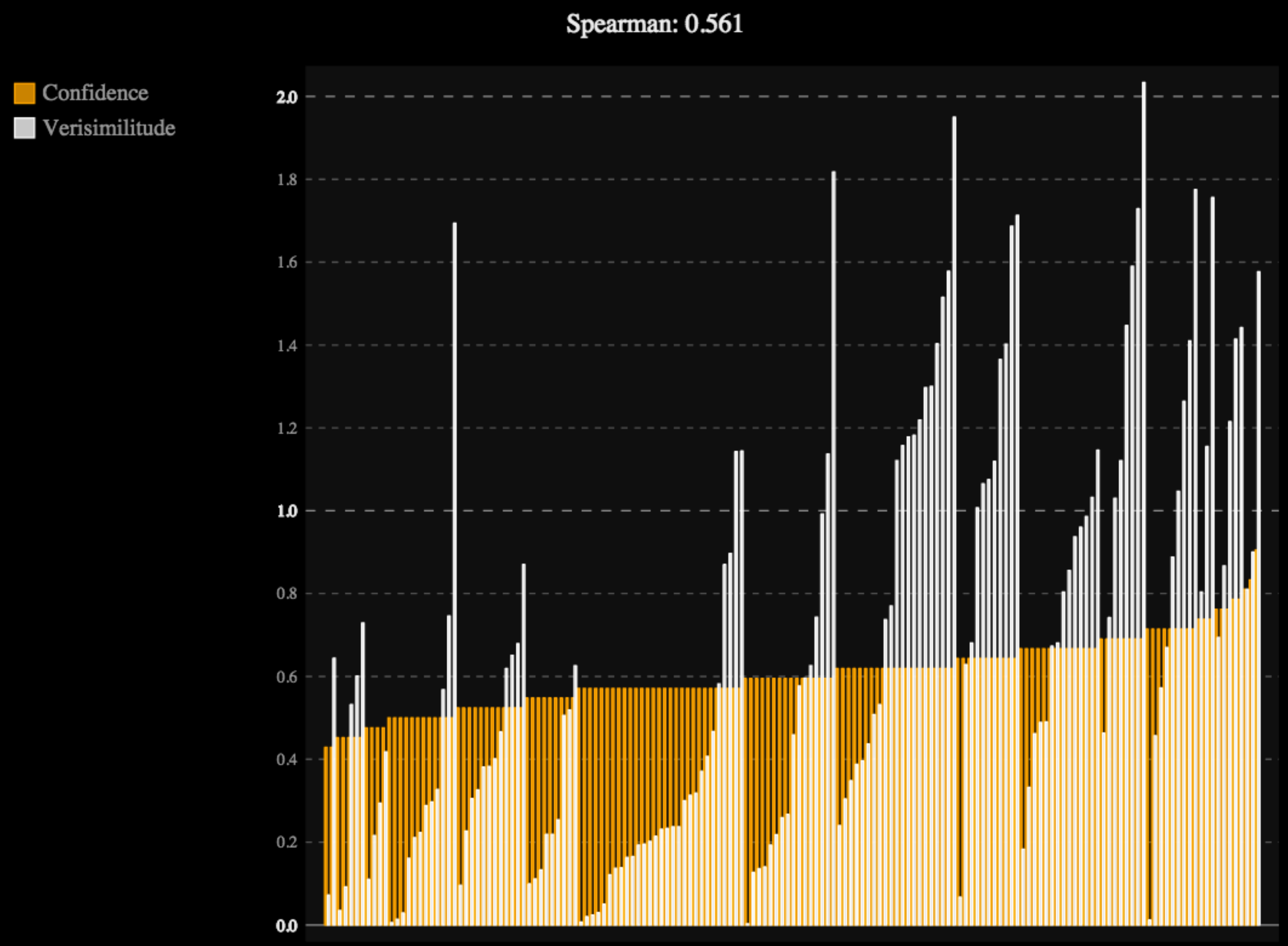
DISTRIBUTION OF SCORES

1m Tweets about Politicians 1-4



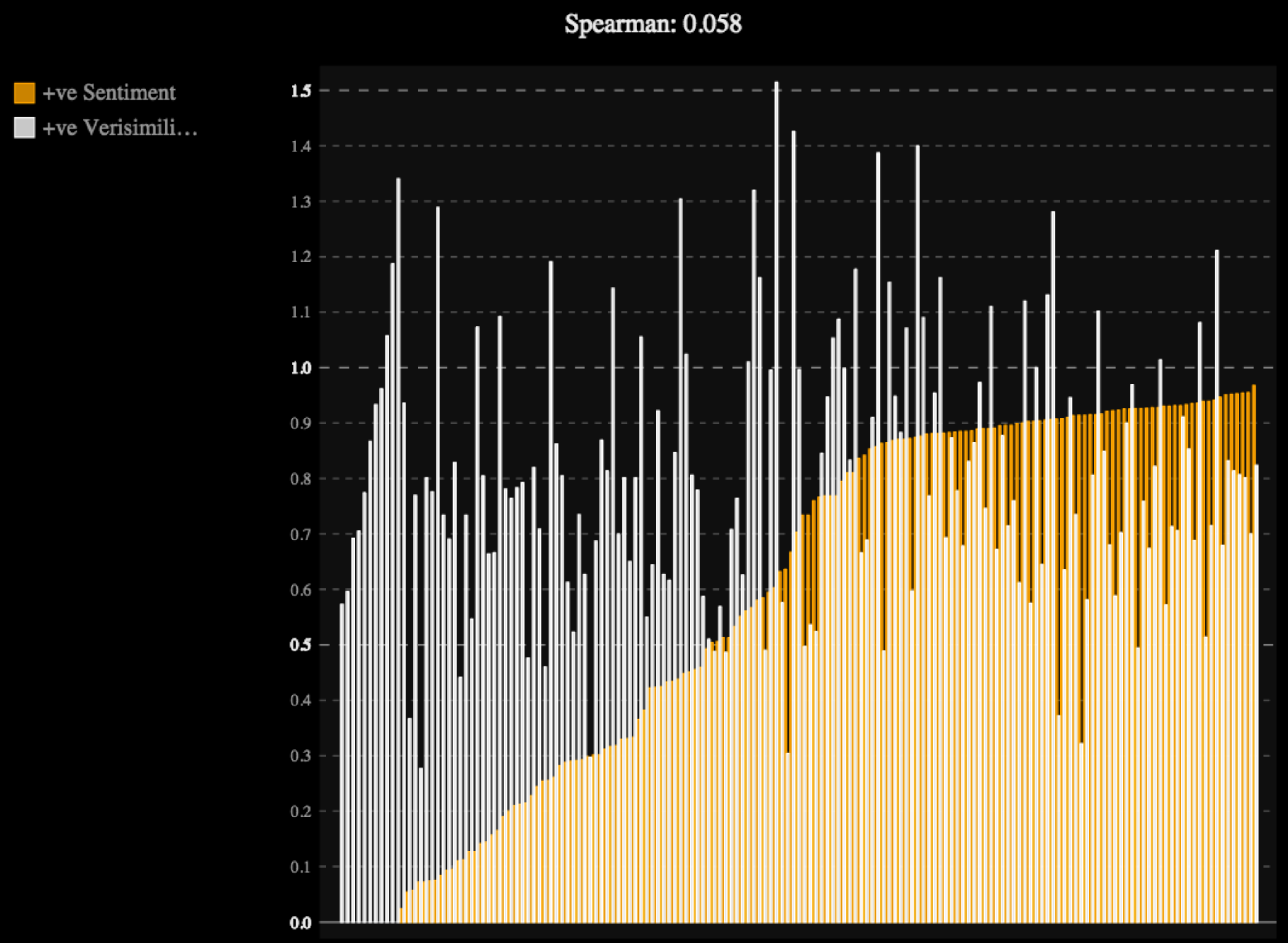
VERISIMILITUDE CONFIDENCE

- Generally, the algorithm has higher confidence in stronger verisimilitude scores



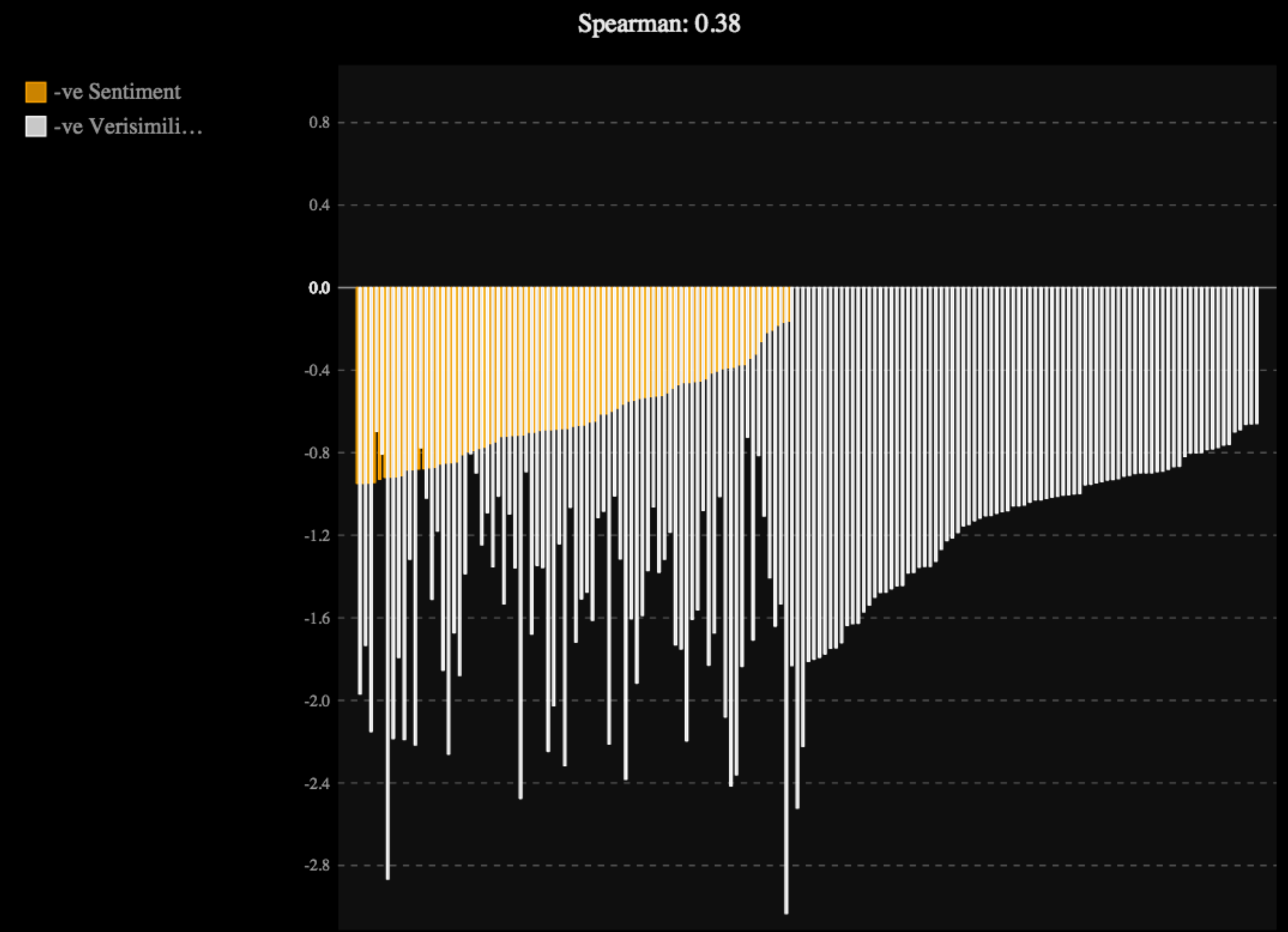
⊕ VERISIMILITUDE : ⊕ SENTIMENT

- Positive verisimilitude and positive sentiment scores exhibit only weak correlation



⊖ VERISIMILITUDE : ⊖ SENTIMENT

- Negative verisimilitude and negative sentiment scores correlate moderately

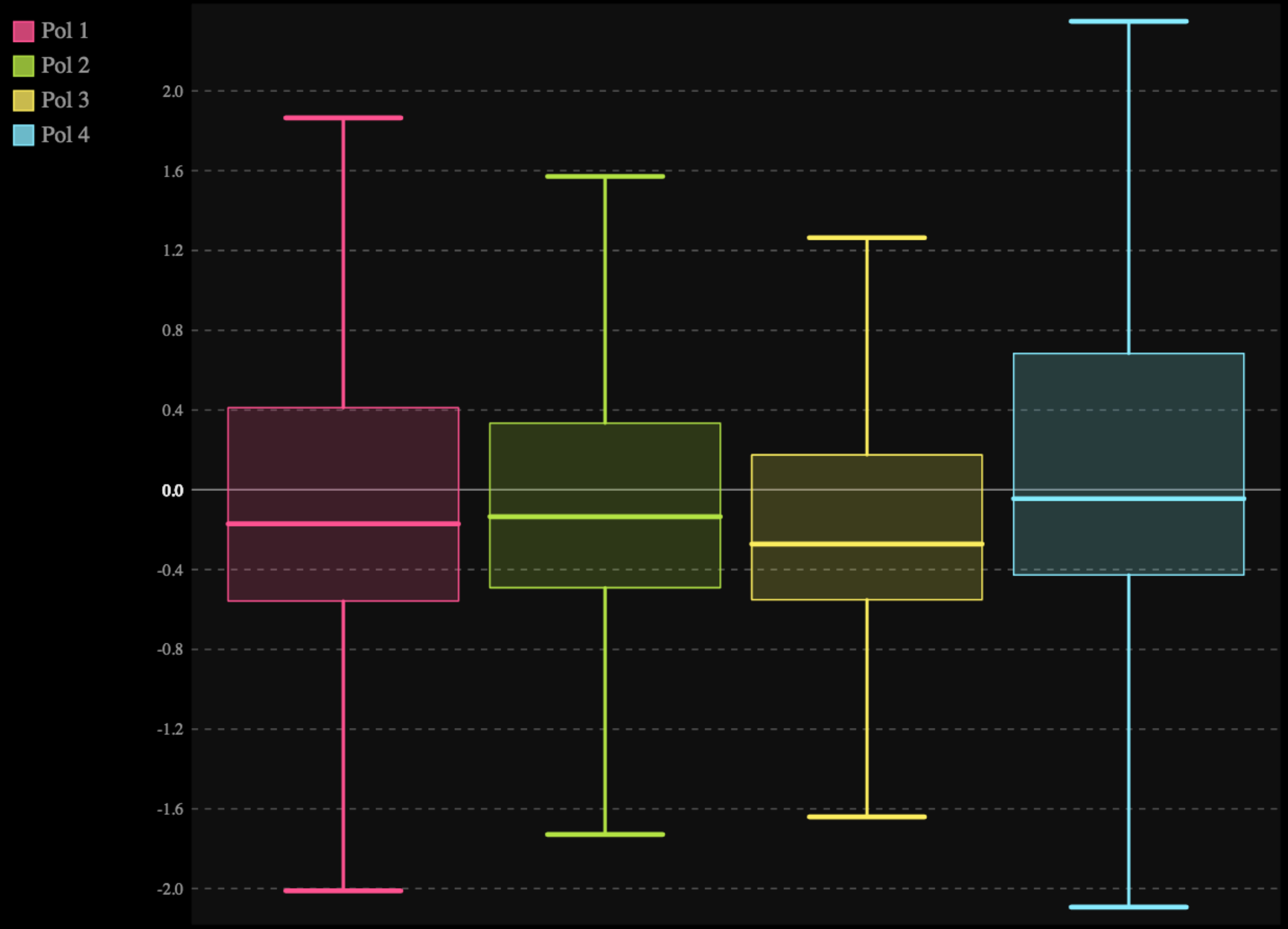


1K TWEETS BY POLITICIANS 1-4

	Sum	Median	Stdev	Min	Max	⊕ %
Pol 1	44.262	-0.171	1.114	-2.593	4.269	41.1
Pol 2	59.829	-0.135	1.218	-2.517	8.106	41.47
Pol 3	-5.917	-0.272	1.056	-2.601	5.017	32.66
Pol 4	316.30	-0.044	1.443	-1.791	12.869	69.38

- All politicians' median values are below zero!
- Politician 4 seems to have the strongest verisimilitude profile (⊕) as well as the greatest dispersion
- Politician 3 has a weak-looking verisimilitude profile (⊖) with the least dispersion

1K TWEETS BY POLITICIANS 1-4

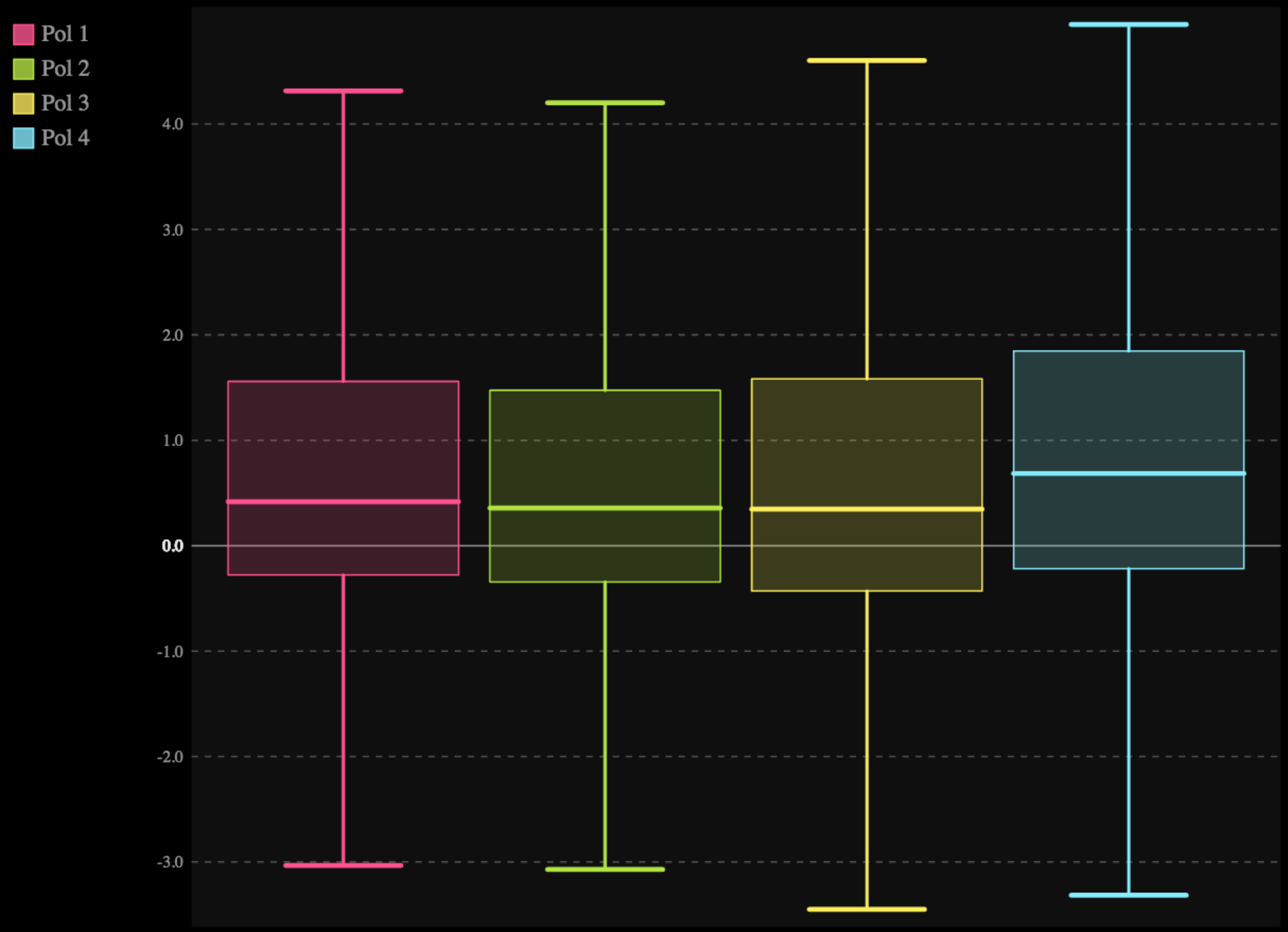


1M TWEETS ABOUT POLITICIANS 1-4

	Sum	Median	Stdev	Min	Max	⊕ %
Pol 1	0.906	0.418	1.756	-2.906	52.729	64.32
Pol 2	0.784	0.357	1.693	-2.723	54.632	62.03
Pol 3	0.875	0.347	1.905	-3.054	35.151	59.85
Pol 4	1.005	0.685	1.656	-3.396	57.902	71.68

- The general public's median values are above zero!
- Politician 4 again seems to have the strongest verisimilitude profile (⊕) but this time the least dispersion
- Politician 3 again has a weaker-looking verisimilitude profile (⊖) with the most dispersion

1M TWEETS ABOUT POLITICIANS 1-4



OUTRO

- We model the qualitative (un)markedness of information in text using a complex composite measure - verisimilitude
- Verisimilitude reflects rational communication principles and maxims pertaining to quantity, quality, relevance, and manner
- Verisimilitude estimation is much more than deception, sentiment, or emotion analysis, and involves a wide range of linguistic devices and extralinguistic phenomena
- Verisimilitude estimates can be used as highly rich and sensitive information filtering and ranking criteria

REFERENCES

• Androutsopoulos, Ion and Malakasiotis, Prodromos (2010). A Survey of Paraphrasing and Textual Entailment Methods. Journal of Artificial Intelligence Research 38.
• Bamman, David and Smith, Noah A. (2015). Contextualized Sarcasm Detection on Twitter. Proceedings of ICWSM 2015.
• Barbieri, Francesco and Saggion, Horacio (2014). Modelling Irony in Twitter. Proceedings of EACL 2014.
• Baumer, Eric et al. (2015). Testing and Comparing Computational Approaches for Identifying the Language of Framing in Political News.
• Castillo, Carlos et al. (2011). Information Credibility on Twitter. Proceedings of WWW 2011.

• Ciampaglia, Giovanni Luca et al. (2015). Computational Fact Checking from Knowledge Networks. PLoS ONE, 10:6.
• Collins-Thompson, Kevyn (2014). Computational Assessment of Text Readability: A Survey of Current and Future Research. International Journal of Applied Linguistics, 165:2.
• Farkas, Richard et al. (2010). The CoNLL-2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text. Proceedings of CoNLL 2010.
• Flor, Michael et al. (2013). Lexical Tightness and Text Complexity. Proceedings of NLP4ITA.

• Ghosh, Aniruddha et al. (2015). SemEval-2015 Task 11: Sentiment Analysis of Figurative Language in Twitter. Proceedings of SemEval 2015.
• Grice, H. P. (1975). Logic and Conversation. In P. Cole and J. L. Morgan (Eds.), Syntax and Semantics 3: Speech Arts. pp. 41-58.

REFERENCES

- Horn, Colby et al. (2014). Learning a Lexical Simplifier Using Wikipedia. Proceedings of ALC 2014.
- Kao, Justine et al. (2013). The Funny Thing About Incongruity: A Computational Model of Humor in Puns. Proceedings of COGSCI 2013.
- Lauri Karttunen and Annie Zaenen (2005). Veridicity.
- Lotan, Amnon et al. (2013).TruthTeller: Annotating Predicate Truth. Proceedings of NAACL-HLT 2013.
- Kate, Rohit et al. (2010). Learning to Predict Readability using Diverse Linguistic Features. Proceedings of COLING 2010.
- Loenneker-Rodman, Birte and Narayanan, Srini (2008). Computational Approaches to Figurative Language.
- Lukin, Stephanie and Walker, Marilyn (2013). Really? Well. Apparently Bootstrapping Improves the Performance of Sarcasm and Nastiness Classifiers for Online Dialogue. Proceedings of LASM 2013.
- de Marneffe, Marie-Catherine et al. (2011). Veridicality and utterance understanding. Proceedings of ICSC2011.
- MacCartney, Bill et al. (2006). Learning to recognize features of valid textual entailments. Proceedings of HLT-NAACL 2006.
- de Marneffe, Marie-Catherine et al. (2008). Finding Contradictions in Text. Proceedings of ACL 2008.

REFERENCES

- Mihalcea, Rada and Pulman, Stephen (2007). Characterizing Humour: An Exploration of Features in Humorous Texts. Proceedings of CICLing 2007.
- Mitra, Tanushree & Gilbert, Eric. (2015). CREDBANK: A Large-Scale Social Media Corpus with Associated Credibility Annotations. Proceedings of ICWSM 2015.
- Moncecchi, Guillermo et al. (2012). Improving Speculative Language Detection using Linguistic Knowledge. Proceedings of ACL 2012.
- Pitler, Emily and Nenkova, Ani (2008). Revisiting Readability: A Unified Framework for Predicting Text Quality. Proceedings of EMNLP 2008.
- Qazvinian, Vahed et al. (2011). Rumor has it: Identifying Misinformation in Microblogs. Proceedings of EMNLP 2011.
- Recasens, Marta et al. (2013). Linguistic Models for Analyzing and Detecting Biased Language. Proceedings of ACL 2013.
- Ritter, Alan et al. (2008). It's a Contradiction—No, it's Not: A Case Study using Functional Relations. Proceedings of EMNLP 2008.
- Reyes, Antonio et al. (2013). A multidimensional approach for detecting irony in Twitter. Lang Resources & Evaluation 47:1.
- Sanchez, Liliana Mamani and Vogel, Carl (2015). A hedging annotation scheme focused on epistemic phrases for informal language. Proceedings of the Workshop on Models for Modality Annotation.
- Roser, Saurí (2008). FactBank 1.0: Annotation Guidelines.

REFERENCES

- Szarvas, György et al. (2012). Cross-Genre and Cross-Domain Detection of Semantic Uncertainty. Computational Linguistics 38:2.
- Shutova, Ekaterina (2015). Design and Evaluation of Metaphor Processing Systems. Computational Linguistics (to appear).
- Zanzotto, Fabio Massimo et al. (2011). Linguistic Redundancy in Twitter. Proceedings of EMNLP 2011.
- Zhang, Renxian and Liu, Naishi. (2014). Recognizing Humor on Twitter. Proceedings of CIKM 2014.

GET THE FACT(OID)S STRAIGHT

TOWARDS DEEPER VERISIMILITUDE ESTIMATION

KARO MOILANEN

TheySay Ltd

Department of Computer Science, University of Oxford

Sentiment Analysis Symposium | 15-JUL-2015