

TOPOI of MATHEMATICAL STATISTICS

Candice Lanius
PhD Candidate
Communication & Media
Rensselaer Polytechnic Institute

Rhetoric Society of America
Atlanta, GA
May 27, 2016

Overview

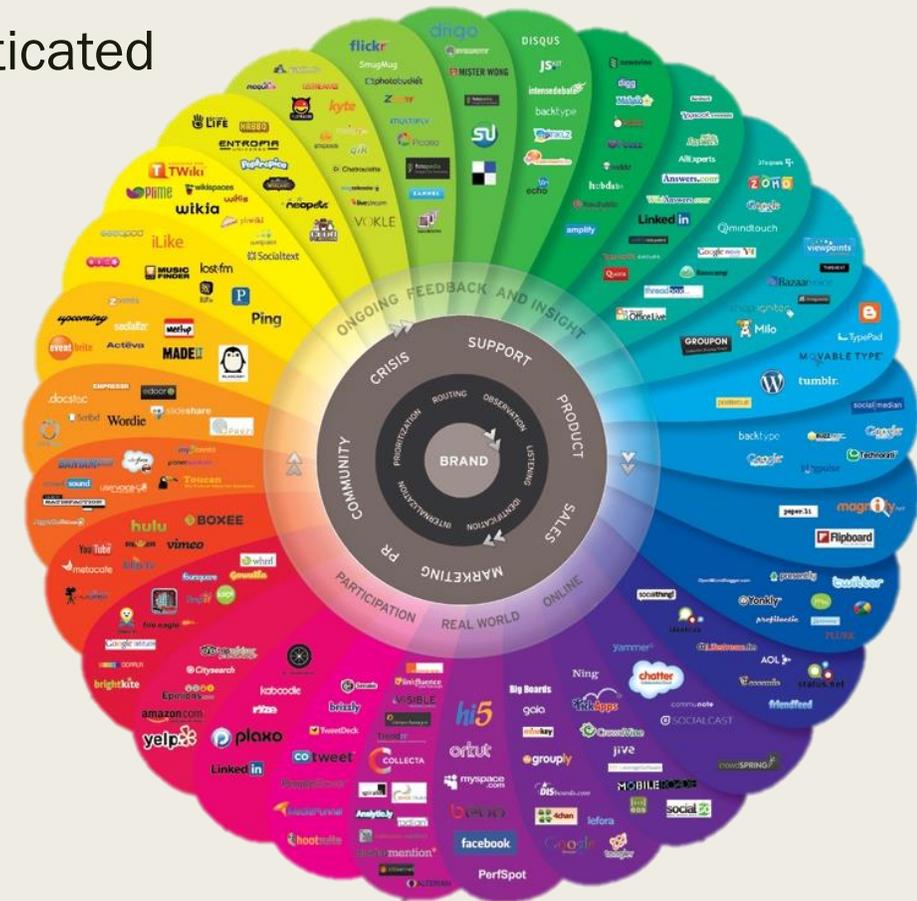
Reveal how aesthetics of mathematical thought—traditionally treated as principles to polish the final report—function as a system of topical invention in the performance of a data analysis.

The image shows a chalkboard with several mathematical equations written in white chalk. The equations are related to probability distributions and calculus. The most prominent equation is the derivative of the natural logarithm of a normal distribution's probability density function with respect to the mean parameter a . Other equations include the expectation of the derivative of the log-likelihood function, the derivative of the log-likelihood function with respect to the parameter θ , and the derivative of the expectation of the log-likelihood function with respect to θ .

$$\frac{\partial}{\partial a} \ln f_{a, \sigma^2}(\xi_1) = \frac{(\xi_1 - a)}{\sigma^2} f_{a, \sigma^2}(\xi_1) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{(\xi_1 - a)^2}{2\sigma^2}\right\}$$
$$\int_{R_n} T(x) \cdot \frac{\partial}{\partial \theta} f(x, \theta) dx = M\left(T(\xi) \cdot \frac{\partial}{\partial \theta} \ln L(\xi, \theta)\right)$$
$$\int_{R_n} T(x) \cdot \left(\frac{\partial}{\partial \theta} \ln L(x, \theta)\right) \cdot f(x, \theta) dx = \int_{R_n} T(x) \cdot \left(\frac{\partial}{\partial \theta} \frac{f(x, \theta)}{f(x, \theta)}\right) \cdot f(x, \theta) dx$$
$$\frac{\partial}{\partial \theta} \int_{R_n} T(x) f(x, \theta) dx = \int_{R_n} \frac{\partial}{\partial \theta} T(x) f(x, \theta) dx = \int_{R_n} \frac{\partial}{\partial \theta} T(x) \frac{f(x, \theta)}{f(x, \theta)} \cdot f(x, \theta) dx$$

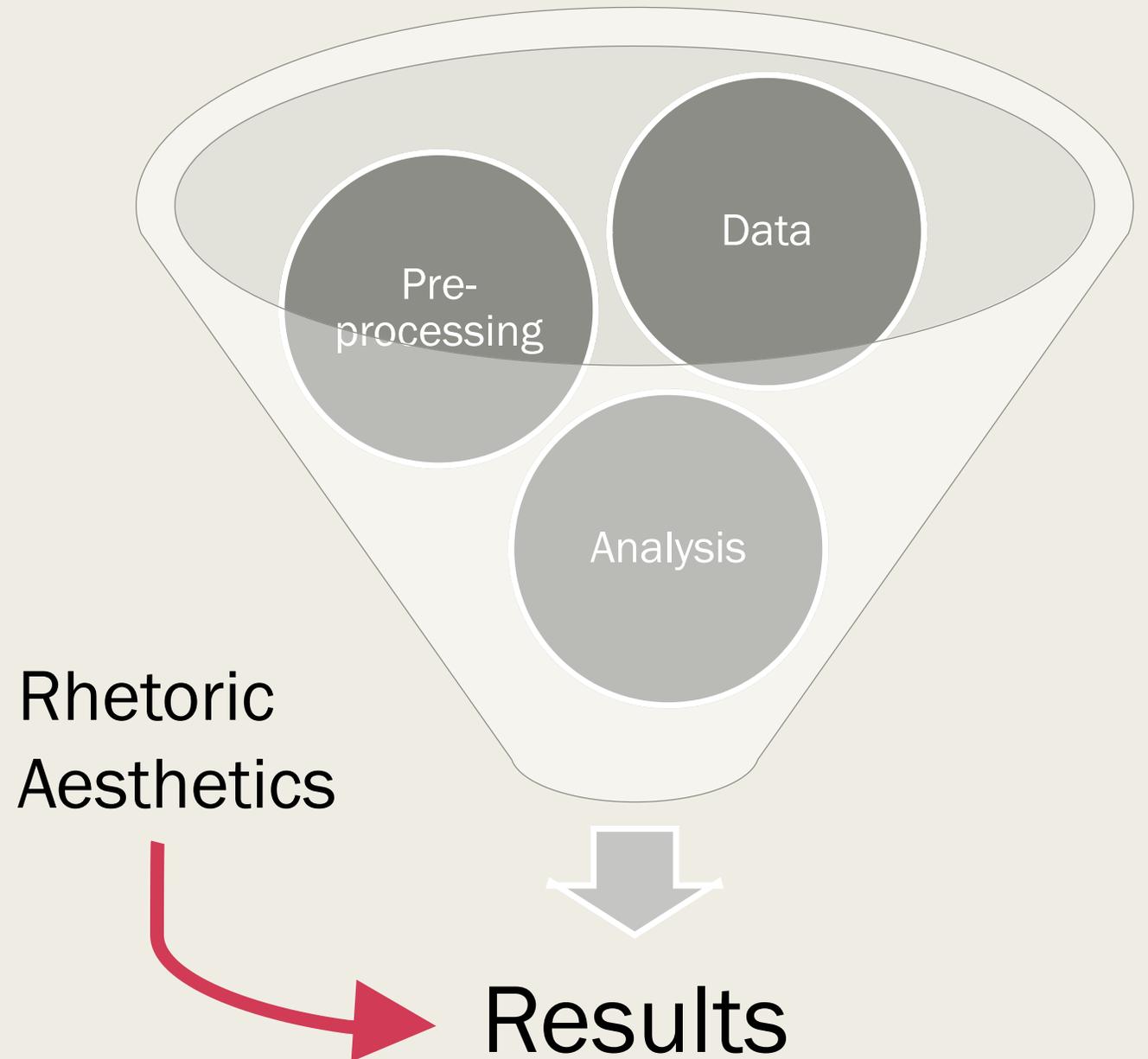
Data Science and Social Media Data

- The “sexy” new realm of mathematical statistics.
- Data mining, machine learning, and other sophisticated computational techniques are applied to social media data collected from information communication technologies.
- Goal to answer important social questions; “extract insights” from the data for:
 - *Marketing and Advertising*
 - *Public and Mental Health*
 - *Economics*
 - *Civic Participation*



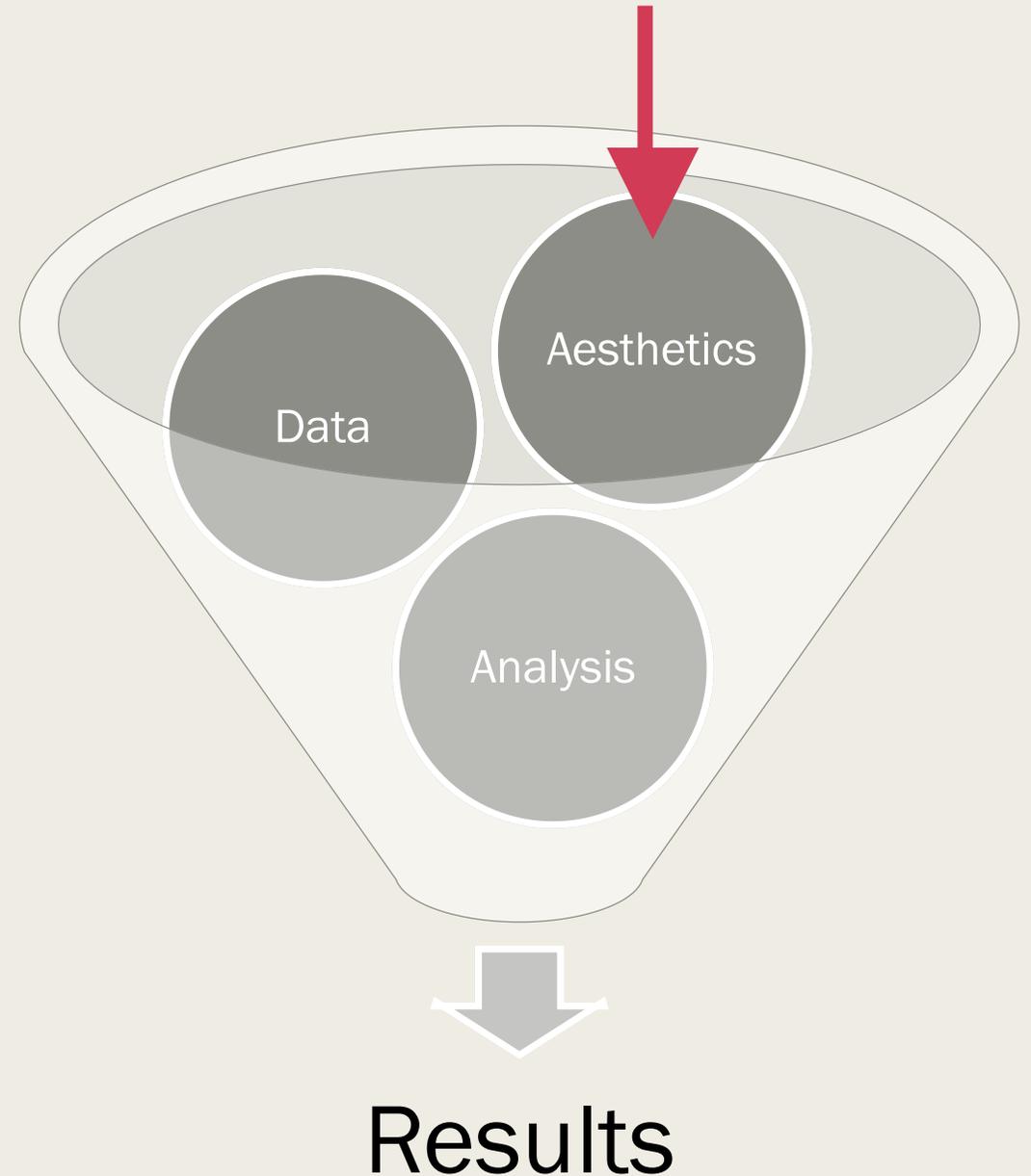
Aesthetics

- Like older notions of rhetoric, aesthetics were traditionally considered a final step to “polish” the results.
- Results come in the form of graphs, data visualizations, and the written report.



Aesthetics as Invention

- I argue that the aesthetic principles of mathematics are a fundamental part of the analytic process.
- Moments where the analyst decides between different options and are not guided by technical features alone.
 - *Data pre-processing, translation, tokenizing, smoothing*
 - *Selection of alpha-criterion for exclusion of results (.05 or .10)*
 - *Theory that relates variables to one another*



Framing: Data Results as Argument or Fact?

Invention

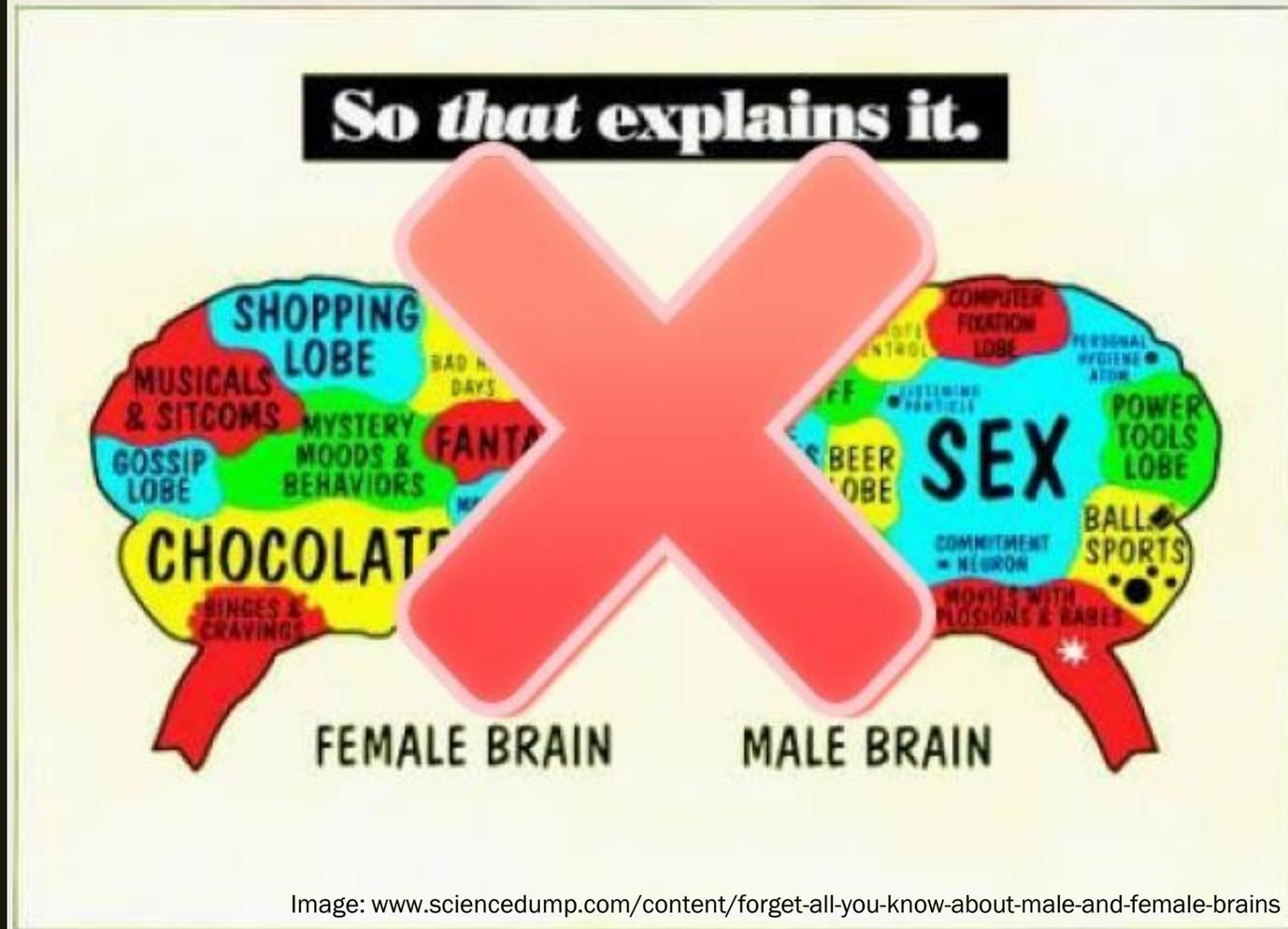
- Discovery of available arguments for a given context
- Process of informed decisions and interpretations
- Demonstration is “persuasive” and “effective.”

Mechanical Objectivity

- Discovery of facts about the world.
- Scientific practice, with systematic mathematical manipulations.
- Results are “valid” and “reliable.”

Data analysis produces *factual arguments* or *argumentative facts*.

- Recognizing this dual nature improves results and facilitates interdisciplinary work.
- Ignoring invention and interpretation leads to the reproduction of biases in research.



Condit, C. (1996). How bad science stays that way: Brain sex, demarcation, and the status of truth in the rhetoric of science. *Rhetoric Society Quarterly*, 26(4), 83-109.

Aesthetic Principles →

Conciseness	Preference for a brief yet complete explanation.
Simplicity	Parsimony- the simpler explanation is preferred.
Clarity	Every component is fully defined and ambiguity is removed.
Structure	Provide a clear division in the data between “signal” and “noise.”
Power	Preference for greater explanatory power of the model.
Cleverness	Showing skill or prowess in analysis.
Surprise	Reveal something about the problem that counteracts existing ideas.

→ System of Topics

Principles Applied in Different Settings

Facilitate Teaching:

- *Sinclair (2004): Mathematical inquiry has aesthetic, cognitive, and affective processes. Aesthetics are an important contribution to teaching mathematical thinking.*

Criticism of Data Visualization:

- *Faltesek (2013): “Mathematical reasoning... can be performed as an identity” (408).*

Data Science

- *My systemization reveals when and where invention occurs within data science to make it apparent and explicit to practitioners.*
- *An explicit system allows practitioners to violate disciplinary norms to respond to new contexts. Better interdisciplinary research.*

Topic 1: Conciseness

Preference for a brief yet complete explanation.

In Big Social Data Analysis, the data precede the research process. Conciseness emerges in the selection of a guiding theoretical framework:

KICKSTARTER

Example: Predict crowdfunding behaviors using the Herding Instinct (e.g. the Bandwagon Effect) (Horvat, Uparna, and Uzzi, 2015).

Topic 2: Simplicity

Parsimony, the simpler model is preferred.

Simplicity emerges in the selection and removal of predictive variables.



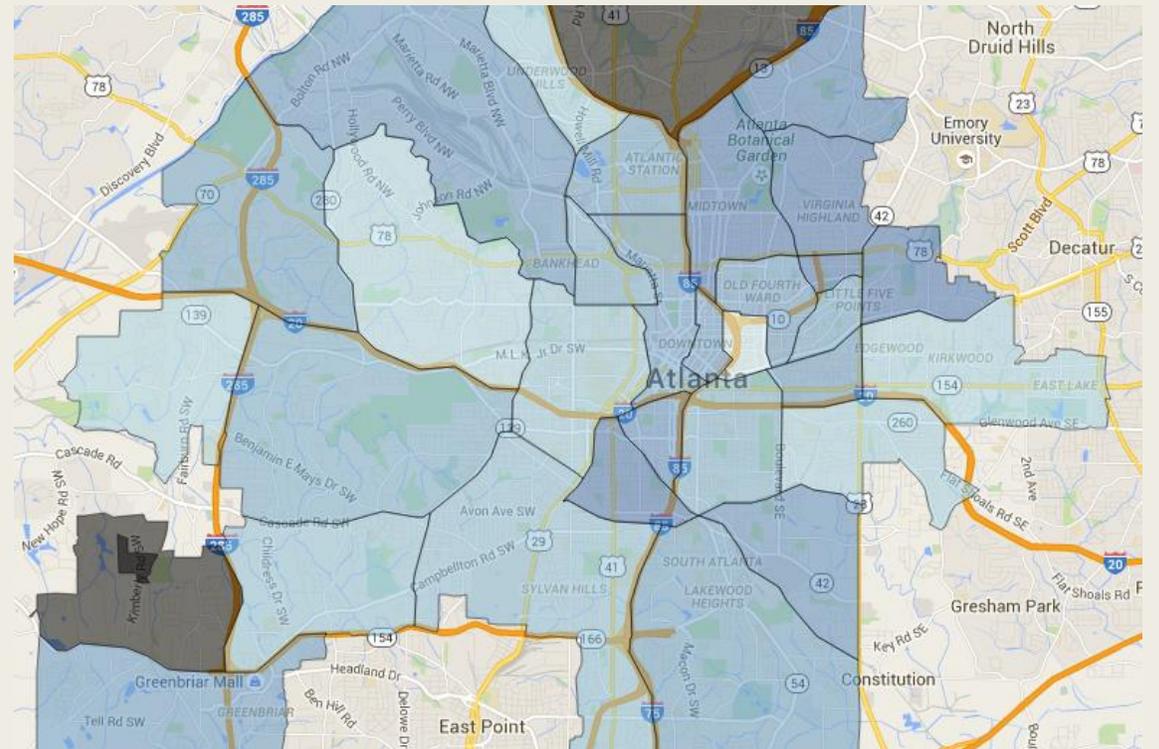
Example: Sentiment analysis of Twitter to predict the stock market uses “positive” and “negative” rather than: calm, happy, angry, or anxious (Bollen, Mao, and Zeng, 2010).

Topic 3: Clarity

Every component is fully defined and ambiguity is removed.

Clarity emerges in the classification of phenomena.

Example: AttitudeBuzz created a library to classify Tweets as homophobic or reclaimed speech (Cohn, Kuntz, and Birnbaum, 2015).



Topic 4: Structure

Provide a clear division in the data between “signal” and “noise.”

Scaffolding to place concepts in relation to each other, where a pattern is meaningful or spurious. Preference for a strong architecture.



Example: Location Based Social Networks, like Four Square, are parsed and analyzed using the principle of *homophily*: shared interests imply social bonds (Kylasa, Kollias, and Grama, 2015).

Topic 5: Power

Preference for greater explanatory power of the model.

Power emerges in the strength of the model to describe or predict the phenomena of interest.

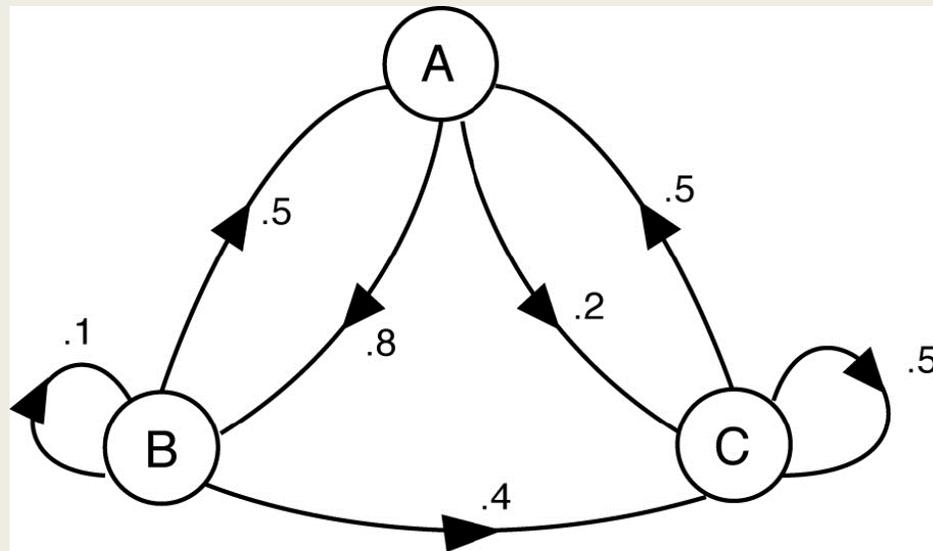
Example: Able to predict post-partum depression using Facebook social interactions and postings from before and after giving birth (De Choudhury, Counts, Horvitz, and Hoff, 2014).



Topic 6: Cleverness

Showing skill or prowess in analysis.

Cleverness, in this field, happens when engineering or hard science methods are applied to social data and sociological questions.



Markov graph of transition probabilities between states A, B and C

Example: Adaptive markov chains to determine the relationship between Twitter and traditional news media for breaking events (Ning, Muthiah, Tandon, and Ramakrishnan, 2015).

Topic 7: Surprise

Reveal something about the problem that counteracts existing ideas.

Surprise emerges in finding something novel or unexpected.



The screenshot shows a Twitter timeline with two tweets from the user Kenneth Cole (@KennethCole). The top tweet is a retweet of a tweet about Egypt, with the text: "Re Egypt tweet: we weren't intending to make light of a serious situation. We understand the sensitivity of this historic moment -KC" posted 1 hour ago. The bottom tweet is a promotional tweet for a spring collection, with the text: "Millions are in uproar in #Cairo. Rumor is they heard our new spring collection is now available online at <http://bit.ly/KCairo> -KC" posted 3 hours ago. The interface includes navigation tabs for Timeline, Favorites, Following, Followers, and Lists.

Example: “Firestorms” on Twitter do not change the pre-existing social structure, so they are unlikely to have biographical or social consequences (Lamba, Malik, and Pfeffer, 2015).

Conclusion

- Treating the principles of mathematical aesthetics as a system of topics helps practitioners understand where they are making interpretive decisions. To put this another way, there are steps with no “right” or “wrong” choice, so the analyst is relying on invention strategies to make persuasive arguments.
- Currently, these moments are beholden to strict disciplinary standards. In the field of big social data, where interdisciplinary research is the norm, being aware of this system of topics helps researchers violate the norms where appropriate.

References

- Condit, C. (1996). How bad science stays that way: Brain sex, demarcation, and the status of truth in the rhetoric of science. *Rhetoric Society Quarterly*, 26(4), 83-109.
- Dreyfus, T., and T. Eisenberg. (1986). On the aesthetics of mathematical thought. *For the learning of mathematics*, 6(1), 2-10,
- Faltesek, D. (2013). Big argumentation? *triple*, 11(2), 402-411.
- Leff, M. (1983). Topical invention and metaphoric interaction. *Southern Communication Journal*, 48(3), 214-229.
- Sinclair, N. (2004). The roles of the aesthetic in mathematical inquiry. *Mathematical Thinking and Learning*, 6(3), 261-284.
- Toulmin, S. (1958). *The uses of argument*. Cambridge, UK: Cambridge University Press.