**Quantifying the Qualitative: Examining Relationships in Literary Texts**

Digital Humanities Panel:

* Relationships, bonds, connections
* Weight (0-1); weighted relationship (thickness of edges)
* Nodes (weight, labeled, categorized)
* Close reading (leveraging computational methods to focus qualitative research)
  + Warrant (patterns, grounds for asking a research question)
  + Lens
* Replication, sharing technology

**Abstract:** How do literary scholars quantify social relationships - which are inherently qualitative - in a text? This question drives the research of all three projects comprising this panel: a social network analysis of Jane Austen’s *Mansfield Park* and *Emma*; visualizing character co-occurrence in Mary Russell Mitford’s *Our Village*; and an algorithmic reading of Richard Wright’s *Native Son*. This panel will discuss the parameters, methods, and limitations of such computational research. Ultimately, we posit that deploying computational methods to guide innovative literary analysis renders this work instrumental scholarship (i.e. work that replicable, referenceable, and extensible).

**Panelists:** Adam J. Kneeland, Lindsey Seatter, and Nadia Timperio

**Introduction (2-2.5 minutes)**

* Research questions?
* Panel topic...why are we coming together as a group?

**Part I (7-7.5 minutes)**

Parameters/raw data - defining what your objects and relationships are

* Nadia: sentiment dictionaries, existing python script, text (“Book One: Fear” and why, critical grounding for that choice)
  + A .txt of “Book One: Fear”; manual line breaks following every end punctuation (except those contained within quotation marks). So essentially, every line is a sentence, or multiple sentences if quoted dialogue.
    - Selected the first book for my limited study because in this part, Wright employs naturalism, which is “the idea that man is constructed by social, economic, and psychological forces beyond his control.” As W. Lawrence Hogue notes in his post-colonial, existential reading of the text, by placing Bigger in various “test-tube situations,” Wright shows how the character has been conditioned to behave in a particular way (Hogue 13).
  + Nodes: Moments when the narrator explicitly mentions the emotions a character is experiencing.
  + Edges: What happens between those identifications of emotion. The journey between emotions, as narrated by an unidentified third-person omniscient, intrusive narrator.
  + Neal Caren's script and sentiment dictionaries …
* Lindsey: excerpted chapters of the text (.txt) (why → characters, social tensions present in the plot), historical understanding of class (Perkins), XML personography
  + Research question: can the social relationships in Austen’s novels be successfully modeled as networks?
  + Importance of making the project manageable within your time and technical competencies
    - Narrowed project from Austen’s entire corpus → entirety of *Emma* → three excerpts from *Emma* and *Mansfield Park*
      * How to choose excerpts? Social tensions in the plot
  + What is a node? “the stuff” (Weingart); objects/ideas you are interested in studying; connecting points between relationships; anchors
    - My project: nodes = characters present in a given chapter
    - Importance of retaining the characters’ dynamic qualities (integral to Austen’s style) → used XML personography to do this in a detailed, yet distilled and formulaic manner
  + What is an edge? “the relationship” (Weingart); connections between “the stuff”
    - My project: interactions between characters
  + What is a weight edge? numerical value associated with the edge/relationship. The thicker the edge = the stronger the bond (traditional use)
    - My project: thicker the edge the greater the traversal of social class/rank boundaries
      * Importance of using weight as a signifier; answer my research question
* **Adam**: print/publishing background of Mitford--short impressions of a rural village, already something that was changing/disappearing, published (and republished) throughout the 19th century in urban periodicals; .txt of *Our Village*, a relatively short collection, easy to analyze as a whole(modular-- short sketches instead of longer chapters); fairly representative of Mitford’s style and subject
  + Nodes-- Verbs and proper nouns (characters, essentially, although if a “national noun” came up it would be significant. Looking for how often named, non-human characters co-occur with human characters, as well as which verbs (mobility/feeling/thinking verbs?)
  + Edges-- co-occurrence within a paragraph. Somewhat imprecise and simplistic, but easier to find/model; a broader starting point. Greater thickness equals more frequent co-occurence which, in Mitford’s model, almost always indicates that Mitford and those two characters interacted in some way.

**Part II (7-7.5 minutes)**

Methods - network analysis, text analysis, python script (algorithmic reading of the text)

* Nadia: editing and running python script → CSV
* Lindsey: CSV, Gephi
  + CSV: comma separated values
  + Two CSVs: nodes (columns: id) → listing the characters; edges (columns: source, target, type, weight) → relationships between characters, their direction, their weight
  + Gephi: open-source network analysis program, import CSVs and *wizzbang!* *(I kinda don’t know how Gephi works so....black boxed?)*
    - Adapt colour, add labels, segregate nodes etc.
* Adam: khcoder
  + Simple to the point that I could do it with a pretty minimal tutorial; a tutorial that actually was step by step (unlike D3 tutorials)
  + Actually worked on my computer (unlike Gephi)
  + meant for text analysis (unlike Cytoscape)
  + Cons: Not easily customized

**Part III (7-7.5 minutes)**

Limitations - scope, black boxing, types of questions you can/can’t answer

* Size of corpus (novice DH scholar = smaller corpus)
  + Talk about breadth versus depth
* Black boxing → red herring question because we all use tech we don’t understand
  + Are close readings black boxed? How do we as literary critics ...
  + Healthy dose of skepticism (Felski)
  + Replication of method → GOAL … however, how do you guarantee when you are black boxed? (discuss the inability to replicate Ramsay’s argument → is this due to our inability to understand his methods? Or did he make a fatal flaw?)
* Material hardware limitations and access issues
  + Barriers in labour
* Element of surprise → Adam’s example of having to integrate verbs and that actually being outside original research but bringing to light interesting results. Maybe unavoidable when arguing with (alongside?) computers?
* Nadia: Troubleshooting with Jentery using the “trial and error” method...
  + There are two ways to run Python: by running the script in IDLE, or in the command line. My script (for some unidentifiable reason) only yields results when it is run through the command line.
  + I have both Python 3 and 2 on my computer. I was not aware initially that these two versions use different languages. One major difference is that Python 3 was unable to read selected words in my text file that reflect in the African American vernacular (ex. Awright, Chrissakes). I needed to ensure that Python 2 was running, and not Python 3, for this reason.

**Conclusion (5-5.5 minutes)**

* Essays as instruments
  + DH = instrumental = “product” (Balsamo)
  + Essays = instrumental = influence/enable/validate further scholarship