What is Fake News?
(Rubin, Chen, Conroy ‘15)

- **Serious Fabrication**
  - Exposes, Well written, often by professional journalists
  - Yellow press / tabloids
    - “Clickbait” – eye-catching headlines, often with no relation to the “story”

- **Hoaxes**
  - Goal is often to provide material that is picked up by regular press

- **Humorous fakes**
  - Satire
What tools can we bring to bear?

- Natural Language Processing
- Network
- Machine Learning

Linguistic Tools
(Conroy, Rubin, Chen ‘15)

- “Bag of Words” models
  - Are there words or combinations of words more commonly used in fake news?
- Deep syntax
  - Parse tree rules for deception detection
- Semantic Analysis
  - Identify “known truth” statements in article
- Discourse analysis
  - Similar to sentiment analysis
Network approaches
(Conroy, Rubin, Chen ‘15)

• Linked data
  – Identify subject-predicate-object statement
  – Find shortest path in knowledge base

• Social network behavior
  – Assume links to (trusted) sources in social media reports

Machine Learning

• Classification
  – Training data?

• What do we need for training data?
  – Truthful and deceptive stories
  – Verifiable ground truth
  – Homogeneity in length, style
  – Appropriate timeframe
How is this done? Fact-Checking

- Human verification of claims
  - Requires expert involvement
  - Reliability of fact-checkers
- Automatic verification?

How is this done? Stylometry Example

- Song Feng, Ritwik Banerjee, and Yejin Choi, *Syntactic Stylometry for Deception Detection* Association for Computational Linguistics (ACL 2012)
- Machine Learning approaches
  - Grover – generative network approach – captures machine-generated fake news
  - Defeated by human-generated text
Network Analysis

- Liang Wu, Giovanni Luca Ciampaglia, Huan Liu
  Mining Misinformation in Social Media: Understanding Its Rampant Spread, Harm, and Intervention
  ICDM 2017 tutorial
  – 16-31, 70-, 101-