New Methods for Story Generation in Text Analytics

M. Shahriar Hossain, Chistopher Andrews, Naren Ramakrishnan, Chris North, Francis Quek
CS@VT
Storytelling scenario

L. Garczarek, N. Ramakrishnan, D. Kumar, R.F. Helm, and M. Potts, Global cross-over points in the genome responses of *Synechocystis sp. PCC 6803*, to dehydration, UV-irradiation, and other stresses, manuscript under submission.

L. Garczarek, N. Ramakrishnan, D. Kumar, R.F. Helm, and M. Potts, Global cross-over points in the genome responses of *Synechocystis sp. PCC 6803*, to dehydration, UV-irradiation, and other stresses, manuscript under submission.


Research issues

• What are the theoretical and algorithmic foundations of storytelling?
  – Redescriptions \([KDD’04, AAAI’05, KDD’06]\)
  – Hammocks and cliques

• How do analysts construct stories?
  – Interactive, steerable algorithms
Hammocks and Cliques

- Hammock: A notion of connectivity

Overlap of terms
Hammock Path

- Hammock path: A series of hammocks

A hammock path with width $w=3$
Clique: A notion of neighborhood

Clique size = 4 and hammock width = 3
Clique Path

- **Clique path**: Navigation through tighter linked communities
Atlantic Storm Data (CIA_05 : FBI_28)

Boris Bugarov (a Russian bio-weapon scientist) was hired by Pyotr Safrygin

Algorithm connects Pieter Dopple with Safrygin, but the link is weak.

Clique size=2, hammock threshold=0.99
Example of a story with very weak connection
Algorithm connects Boris Bugarov (a Russian bio-weapon scientist) and Pyotr Safrygin via an intercepted phone call.

Algorithm connects these two documents based on some names of places (connection is vague).

Algorithm finds Pieter Dopple in stone business and money laundering.

Clique size=2, hammock threshold=0.96

Example of a story with stricter connection
Atlantic Storm Data (CIA_05 : FBI_28)

Boris Bugarov (a Russian bio-weapon scientist) was hired by Pyotr Safrygin

Algorithm connects Pieter Dopple with Safrygin, but the link is weak.

Clique size=2, hammock threshold=0.99
Example of a story with very weak connection
Boris Bugarov (a Russian bio-weapon scientist) was hired by Pyotr Safrygin.

Algorithm connects Pieter Dopple with the diamond transactions. Pieter Dopple has relationships with militant Islamic groups.

Clique size=3, hammock threshold=0.99

Example of a better story with small amount of surrounding evidence.
Boris Bugarov (a Russian bio-weapon scientist) was hired by Pyotr Safrygin. This clique explains some intercepted phone calls that involve Pyotr Safrygin who is currently the director of security for the Central Russia Airlines. The conversations were transaction related and involve tanzanite. Algorithm connects Pieter Dopple with the diamond transactions with neighboring evidences. Pieter Dopple has relationships with militant Islamic groups.

Clique size=4, hammock threshold=0.95, Example of a better story with more surrounding evidence
Details

• How do we find stories without materializing the entire similarity network?
• Can we identify landmarks or “hot spots” in stories?
• How can analysts steer the story construction process?
Redescription Mining

The value in cell $(t_x, d_j)$ indicates the frequency of term $X$ in document $j$. 
How do analysts construct stories?

• Incremental progressions
• Organize into clusters, weave story through clusters
• Scatter supporting evidence (story fragments) and compose into a story
User-guided storytelling

- **Spatial cues to guide storytelling**
  - Analysts use the space in meaningful, semantic ways
  - We would like to exploit that computationally to guide the storytelling algorithm
  - Provide feedback in the context of the user’s spatial structures
Demo

• Can mix and match algorithmic and visual interaction capabilities

• Working on user-driven feedback from visualization into storytelling algorithm
Questions?