
New Methods for Story Generation in Text Analytics

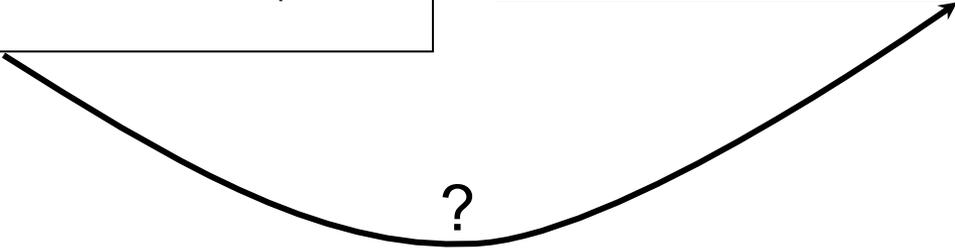
M. Shahriar Hossain, Christopher Andrews,
Naren Ramakrishnan, Chris North, Francis Quek
CS@VT



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.

M.B. Roth and T. Nystul, Buying time in suspended animation, *Scientific American*, Vol. 292, No. 6, pages 48-55, June 2005.



?

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.

CBS domains

L. Schmitt and R. Tampe, Structure and mechanism of ABC transporters, *Current Opinion in Structural Biology*, Vol. 14, No. 4, pages 426-431, Aug 2004.

**Molecular complexes
of CBS domains**

M.B. Roth and T. Nystul, Buying time in suspended animation, *Scientific American*, Vol. 292, No. 6, pages 48-55, June 2005.

Hydrogen sulfide

C. Tang, X. Li and J. Du, Hydrogen sulfide as a new endogenous gaseous transmitter in the cardiovascular system, *Current Vascular Pharmacology*, Vol. 4, No. 1, pages 17-22, Jan 2006.

**Ligands bound
to CBS domains**

J.W. Scott, S.A. Hawley, K.A. Green, M. Anis, G. Stewart, G.A. Scullion, D.G. Norman, and D.G. Hardie, CBS domains form energy-sensing modules whose binding of adenosine ligands is disrupted by disease mutations, *Journal of Clinical Investigation*, Vol. 113, No. 2, pages 182-184, Jan 2004.

DIVERSE DISCIPLINES, ONE COMMUNITY

BiomedicalComputation
REVIEW

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
-

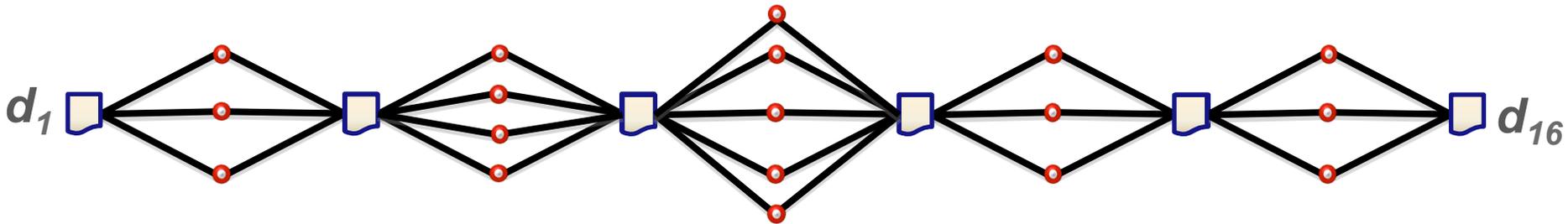
Hammocks and Cliques

- **Hammock** : A notion of connectivity



Hammock Path

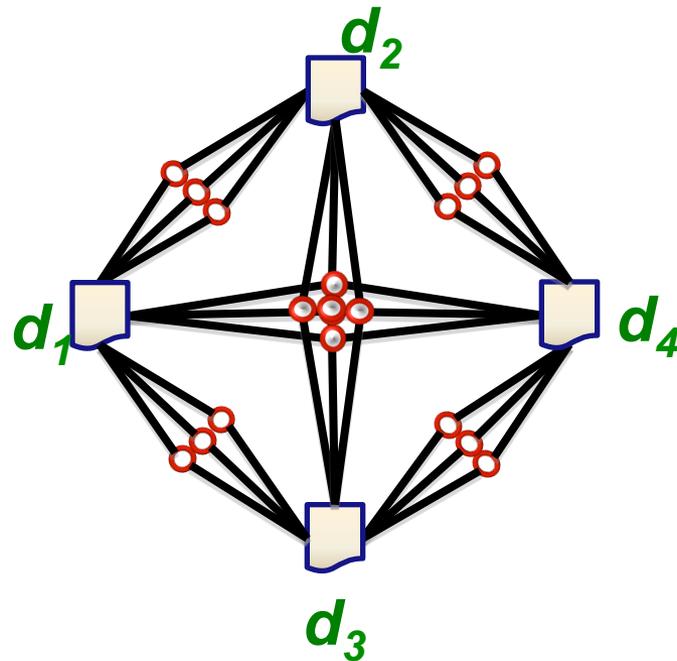
- **Hammock path:** A series of hammocks



A **hammock path** with width $w=3$

Cliques

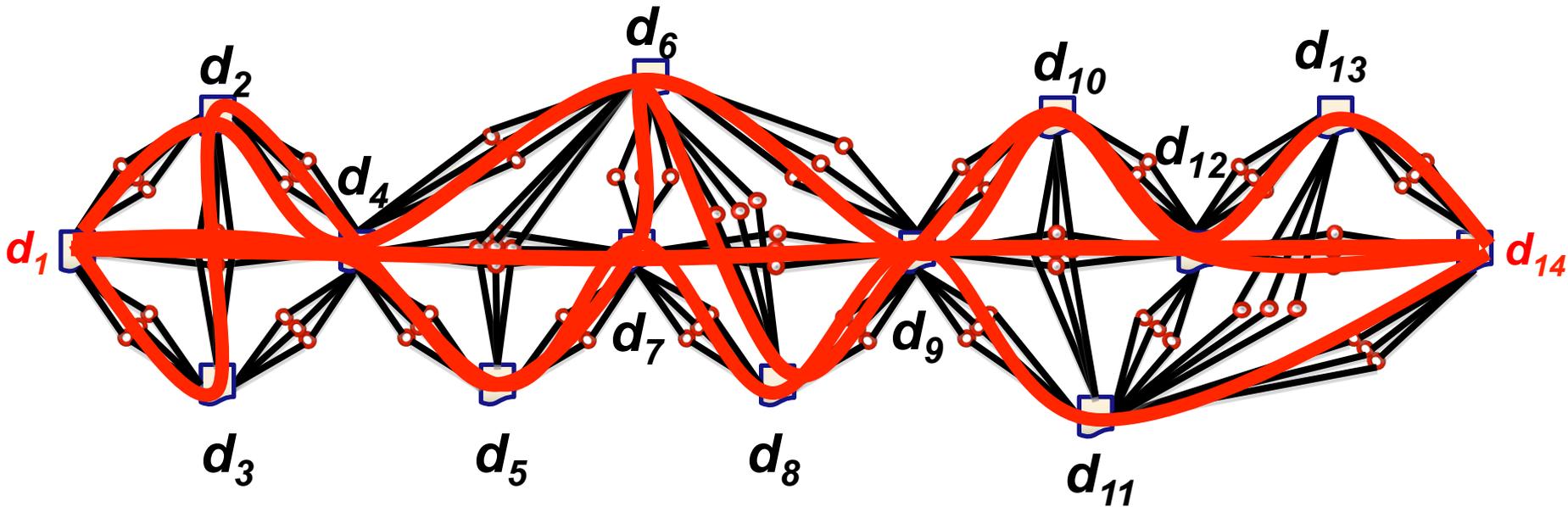
- **Clique:** A notion of neighborhood



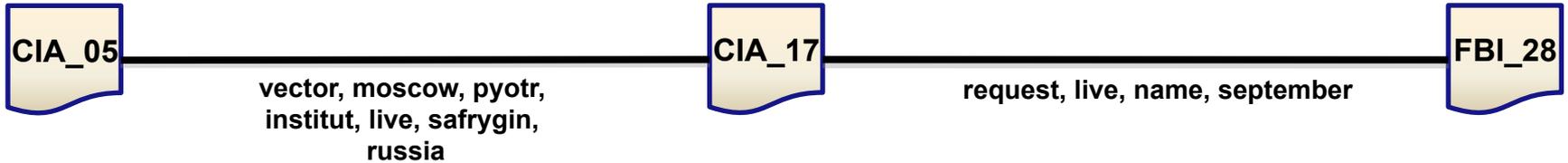
Clique size= 4 and *hammock width=3*

Clique Path

- **Clique path:** Navigation through tighter linked communities



Atlantic Storm Data (CIA_05 : FBI_28)

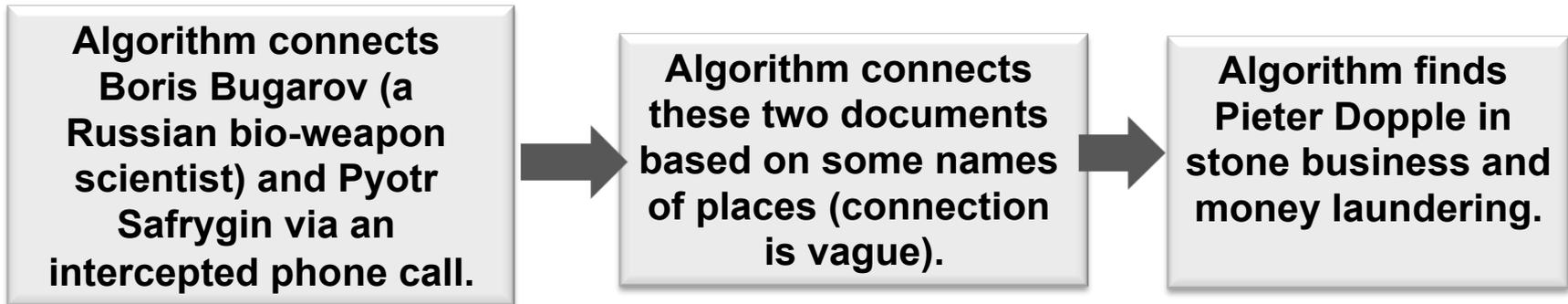
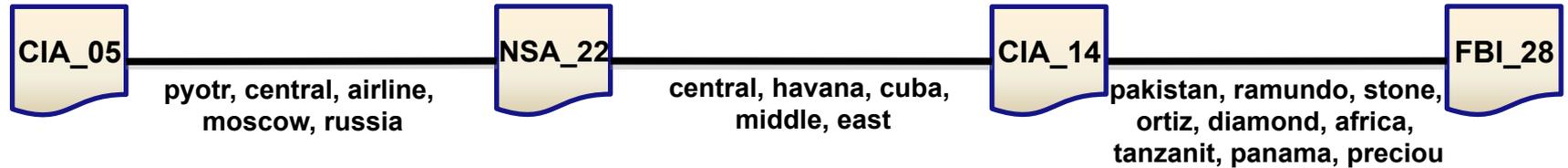


Boris Bugarov (a Russian bio-weapon scientist) was hired by Pyotr Safrygin

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

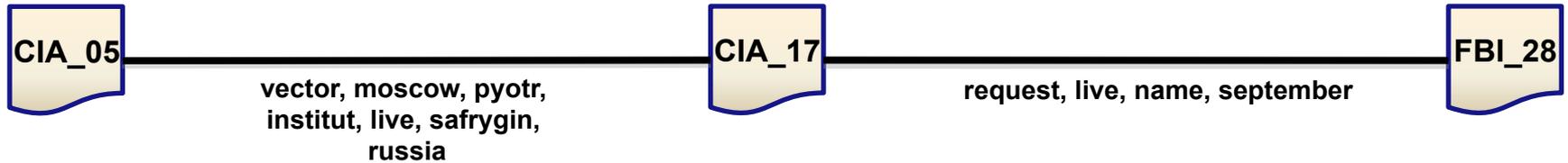
Atlantic Storm Data (CIA_05 : FBI_28)



Clique size=2, hammock threshold=0.96

Example of a story with stricter connection

Atlantic Storm Data (CIA_05 : FBI_28)

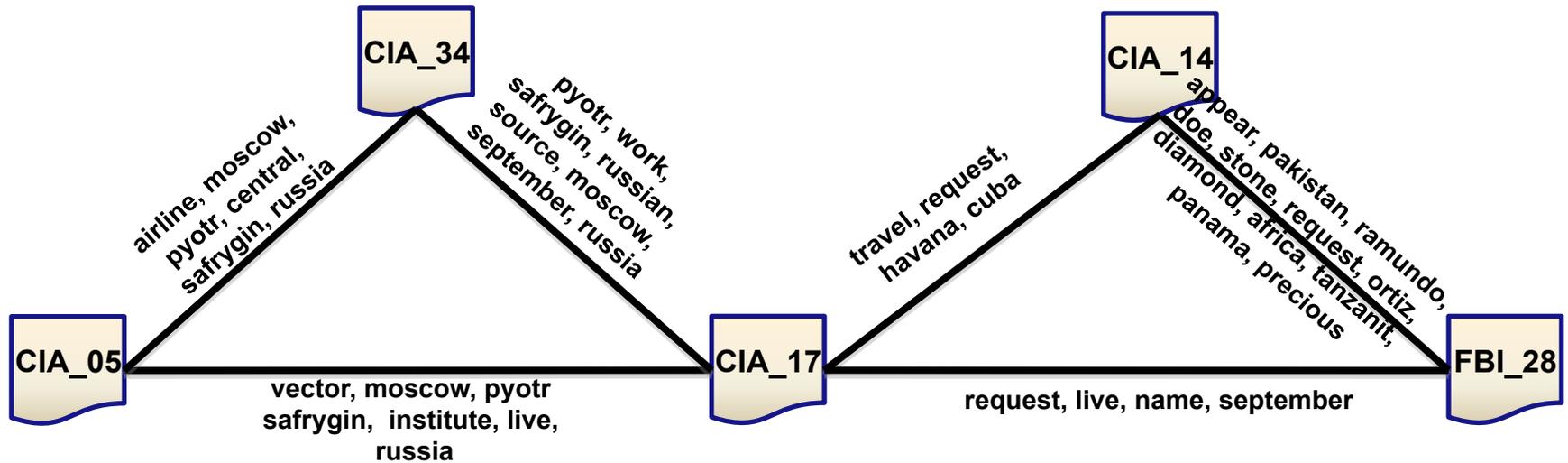


Boris Bugarov (a Russian bio-weapon scientist) was hired by Pyotr Safrygin

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

Atlantic Storm Data (CIA_05 : FBI_28)



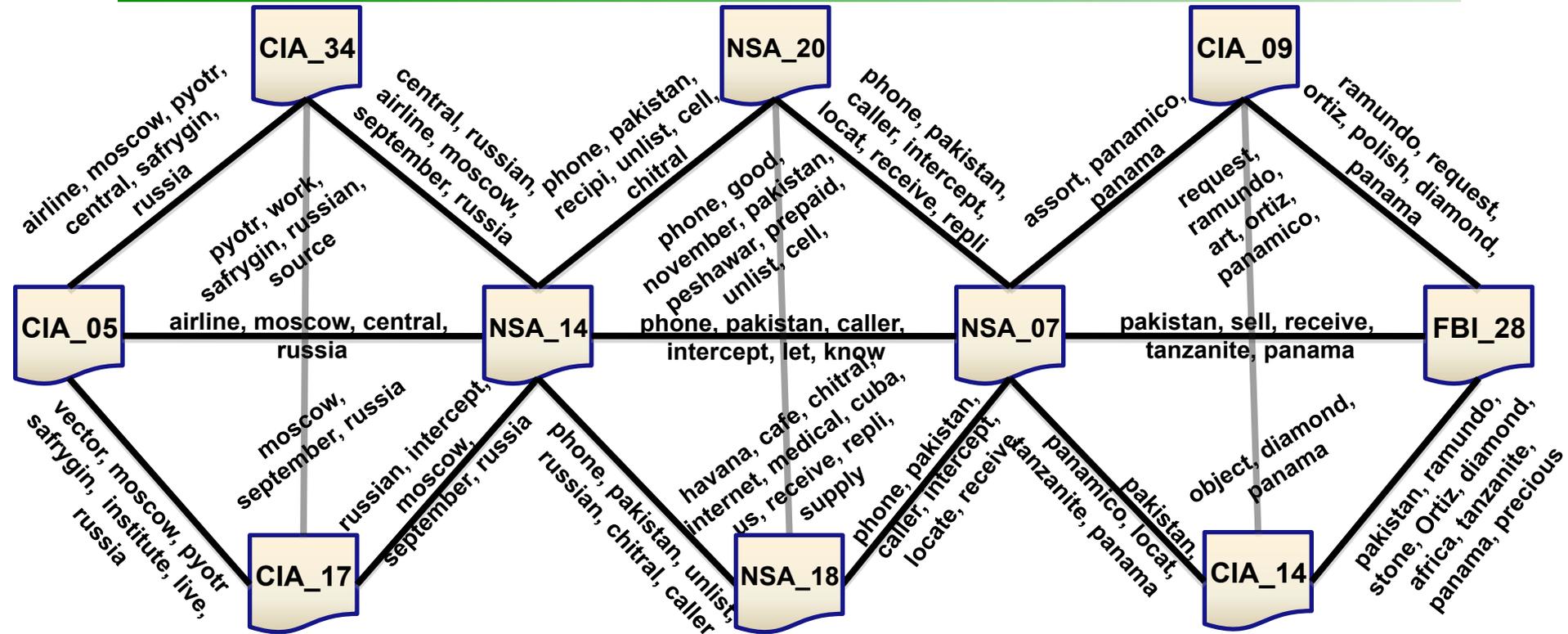
Boris Bugarov (a Russian bio-weapon scientist) was hired by Pyotr Safrygin

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

Atlantic Storm Data (CIA_05 : FBI_28)



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

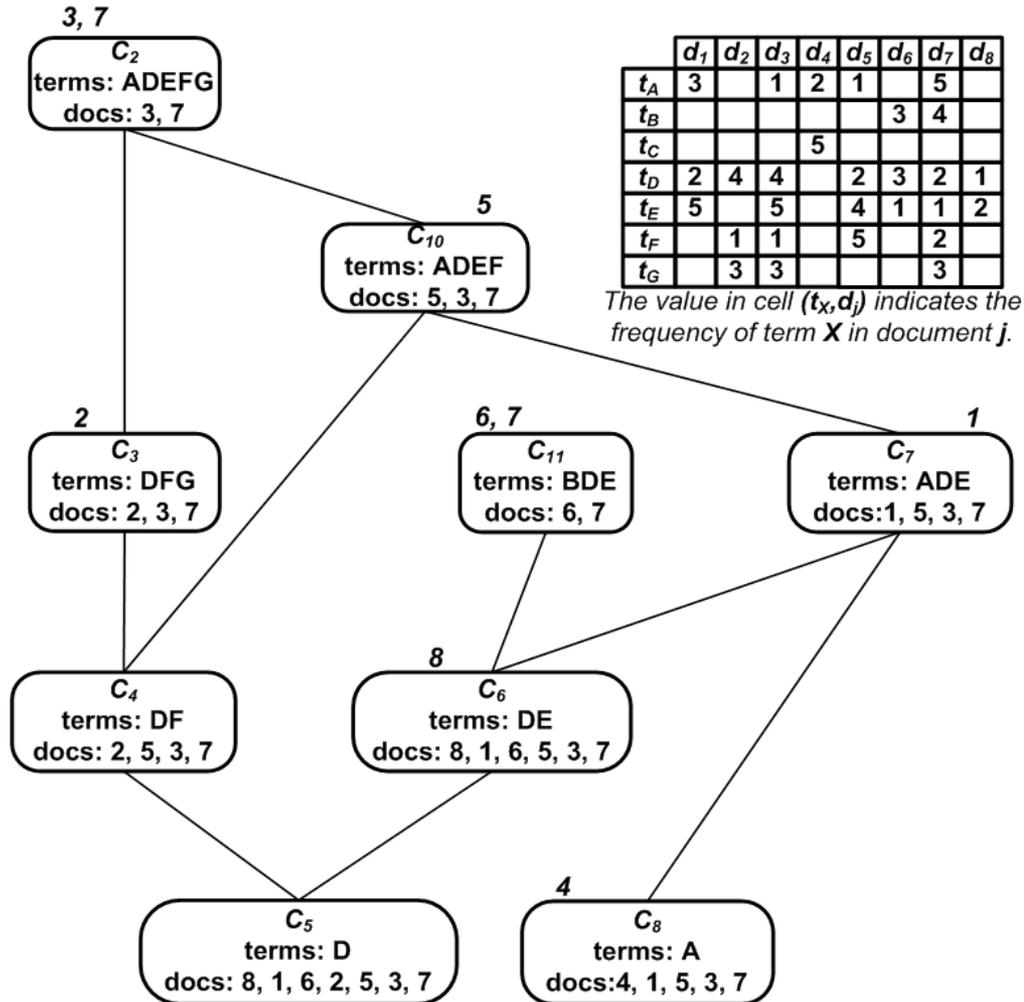
This clique explains some intercepted phone calls that involve Pyotr Safrygin who is currently the director of security for the Central Russia Airlines. The conversations were transaction related and involves tanzanite.

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

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

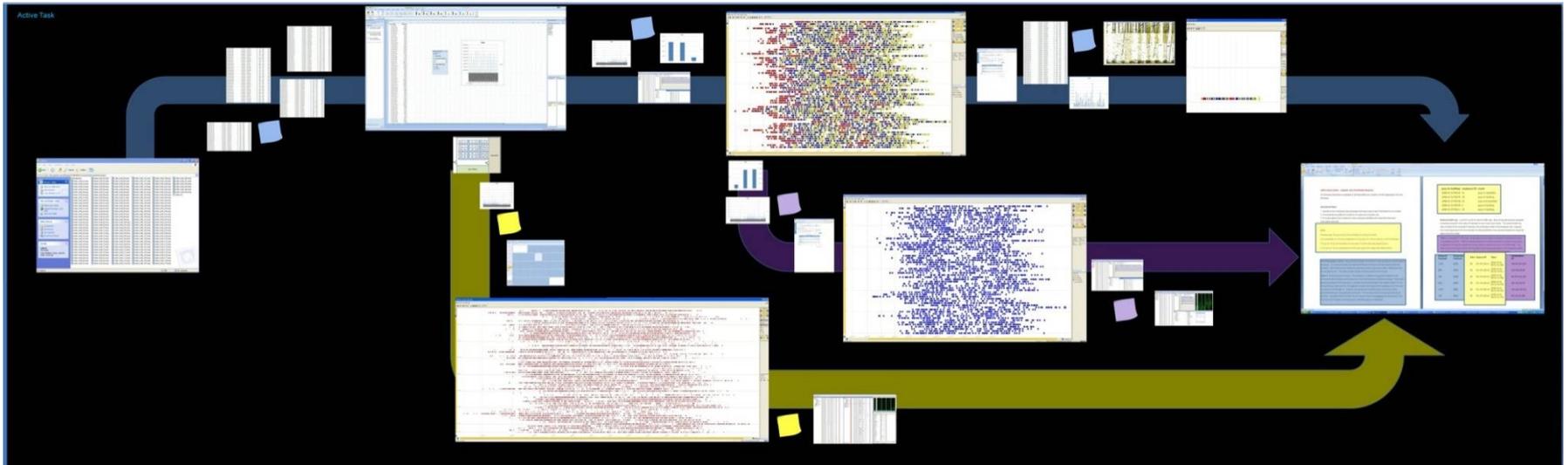


How do analysts construct stories?

- 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?

