Modeling Activist Discourse on State Repression Through Text Analytics

Authors:
Evan Brown
Hailey Reeves

Project Description

- We analyzed trends in activist discourse on state repression in recent decades, using a data set of 94 documents related to state repression.
- We define state repression as the actual or threatened use of penalization against a person or an organization committing nonviolent dissent within state jurisdiction.
- Our data set is a corpus of texts that were gathered by Heidi Reynolds-Stenson, a PhD candidate within the University of Arizona's Department of Sociology. The texts were referenced in interviews she held with over forty Arizonan activists who experienced state repression.

Scientific Challenges

- Qualitative information, such as perception, can be difficult to gauge across multiple narratives. Perception has a direct connection to why and how people engage with causes they deem important. Our analysis will determine underlying themes that influence activist dis/engagement as well as common perceptions held by activists within the sample set.

Potential Applications and Impact

- Data and trends will be utilized in Reynolds-Stenson's dissertation.
- Future directions include analyzing patterns in how activists discuss various types of state repression, as well as what emotions are associated with each repression type.
- A larger sample set across a broader range of social movements could determine whether the same perceptions regarding repression are present in a wider demographic of activists across a greater span of time.

Methodology

1. The documents were processed through optical character recognition software. Frequency tables were generated through R, and a term-document matrix was created.
2. Content similarity was determined through latent semantic analysis (LSA), using a cosine similarity measure.
3. A network was created through NetworkX based upon high frequency terms. Measurements, such as centrality, clustering, and degree distribution, were generated.
4. Similar term information was gathered for the highest-degree terms. Edge weight was defined by their cosine similarity measure. Each term-group was a subgraph; the final graph was the union. Maximal matching, clustering, and degree were calculated.

Results

The most connected and clustered terms illustrate a narrative of conflict that parallel Reynold-Stenson's findings. These terms and their latent semantic connections indicate layers of opposition, most significantly with public street demonstrations and the police. Consequential terms such as "arrested", "convictions", "court", "force", "law", and "rights" had high centrality measures relative to the rest of the network (>0.14). This provides context to the conflict between the two most high degree terms, "police" and "people." Maximal matching pairs include ("control", "enforcement") and ("surveillance", "people").

Definitions

Betweenness Centrality: The sum of the fraction of all-pairs shortest paths that pass through a node.

$$c_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

Clustering Coefficient: A percentage that measures how well a group of nodes are cliqued, from the reference point of an individual node.

$$C_i = \frac{2n_i}{k_i(k_i-1)}$$

Degree Centrality: The number of edges or connections incident on a vertex.
Maximal Match: A subgraph is a maximal match of a larger graph if every edge in the larger graph has a non-empty intersection with at least one edge in the subgraph.
Latent Semantic Analysis: Manipulating a term-document matrix using SVD and a low-rank approximation of the resultant matrices in order to extract semantic information and to better see similarities between terms and documents.
Term-Document Matrix: A matrix constructed from a corpus of documents, with each entry corresponding to the frequency of a term in a document.
Singular Value Decomposition (SVD): The process of splitting a matrix into the product of three matrices, according to the equation $\mathbf{M} = \mathbf{U\Sigma V^*}$ where the middle matrix is diagonal, containing only the matrix M's singular values as entries.

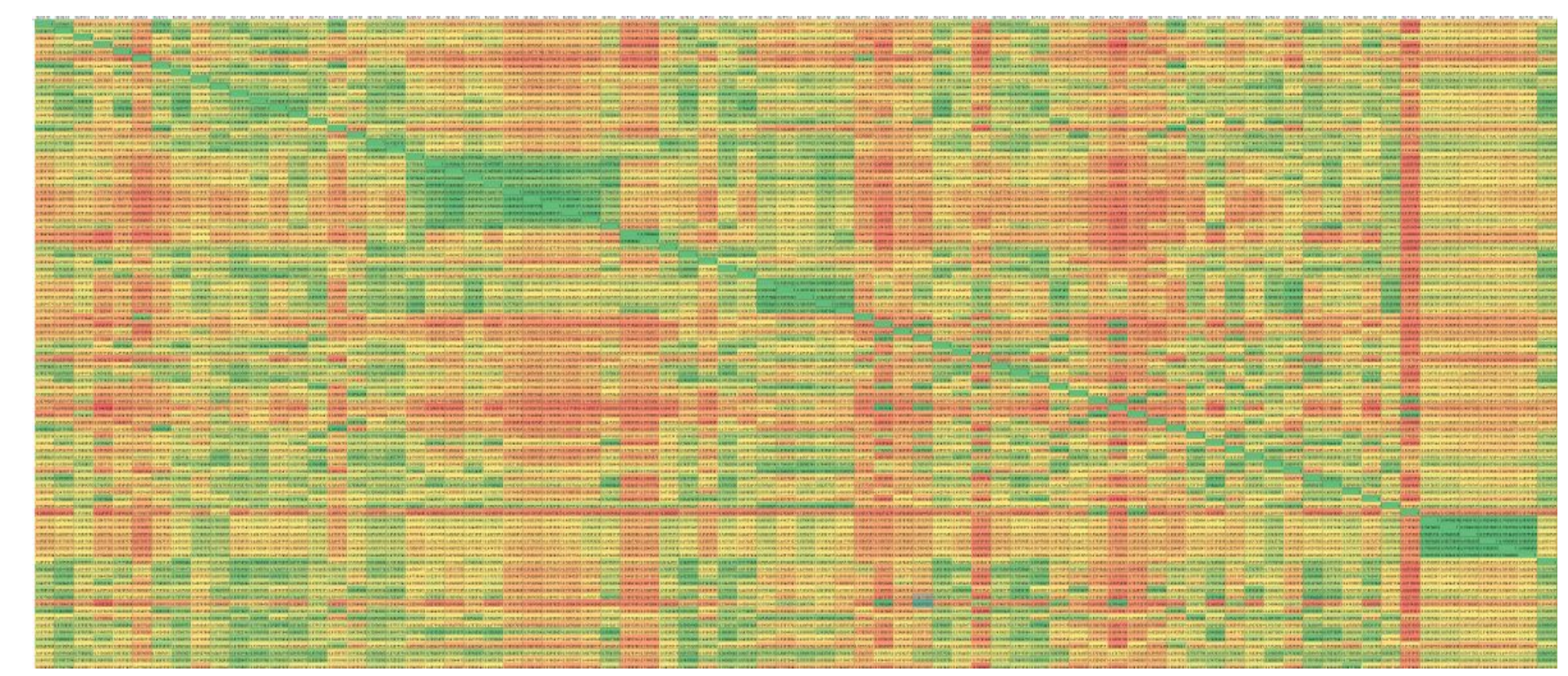

Figure 4. The similarity matrix of all 94 document vectors post-LSA, colored on a red-green gradient for cosine values of 0 to 1. Green entries indicate higher similarity between documents.

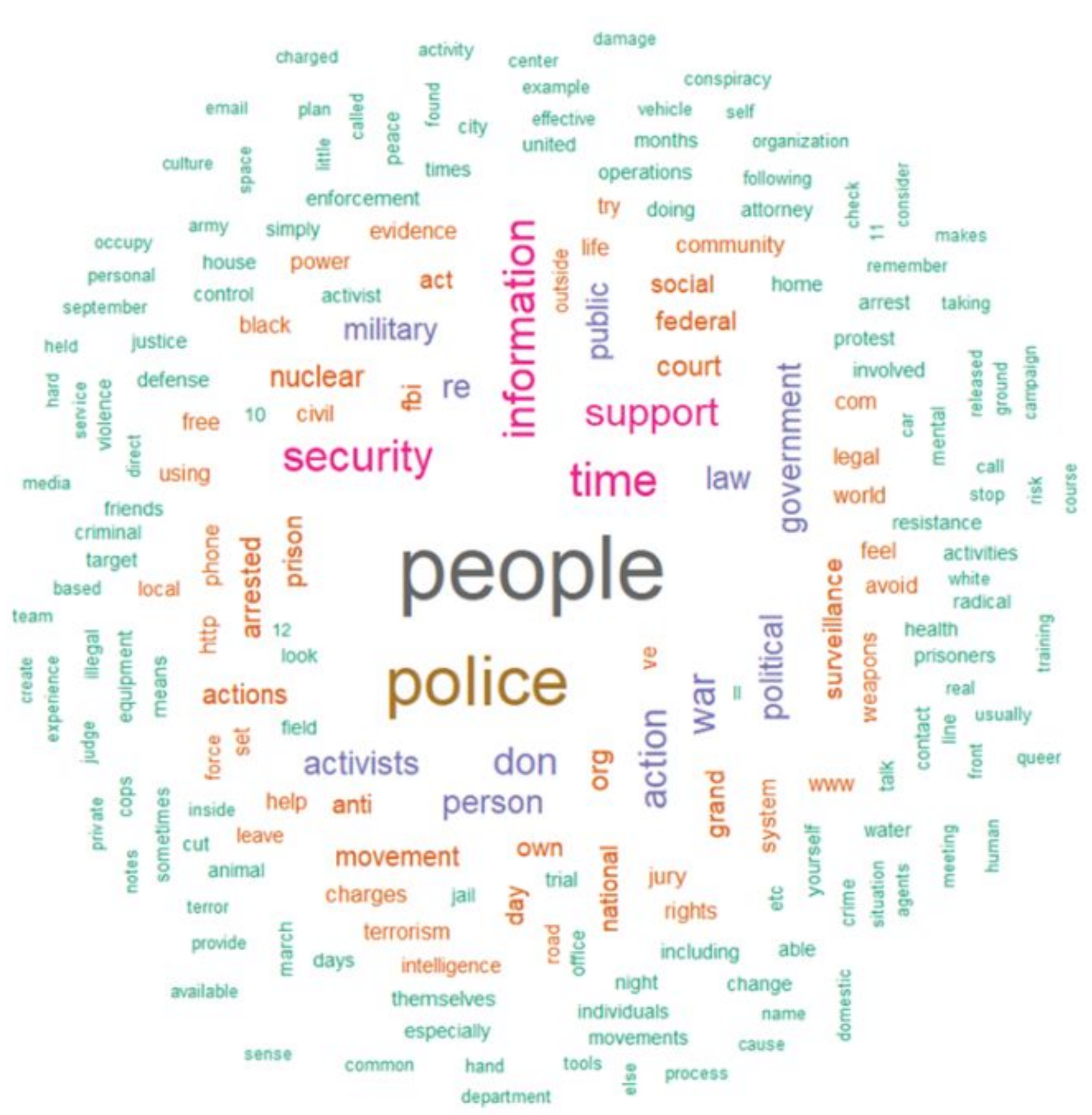

Figure 1. A word cloud generated by R, corresponding to the most frequent terms across all documents, which also corresponds to the "frequency network" manually generated.

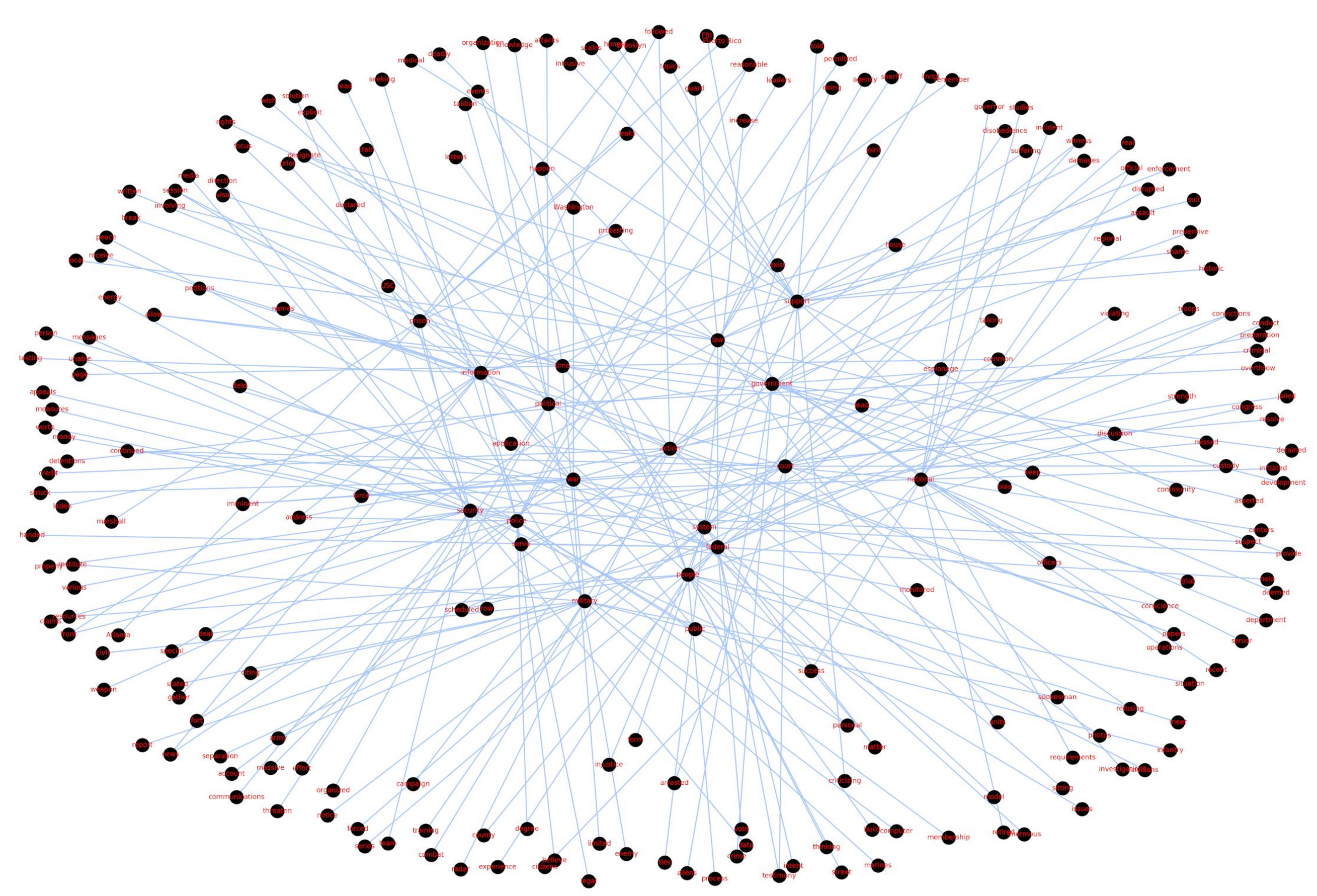

Figure 2. Frequency network post-LSA, displaying high degree nodes within original frequency network and their corresponding LSA information. Significant terms were determined based upon their degree centrality from the original network.

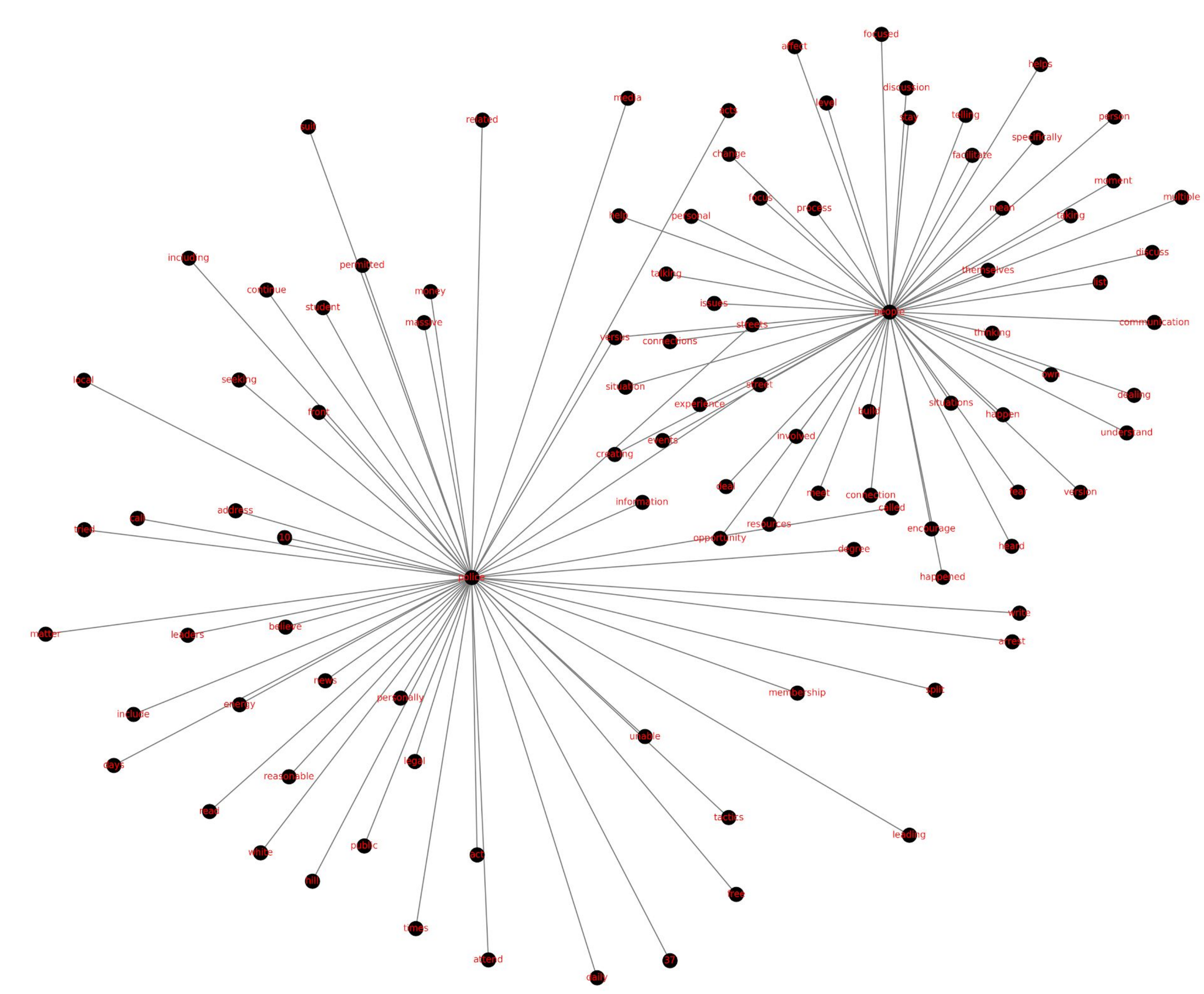

Figure 3. Two networks comparing LSA information with the terms "people" and "police". The networks intersect at the vertices "street" and "versus".

References

1. Anand, Ashish. "Complex Network Theory: An Introductory Tutorial." Department of Computer Science and Engineering. Indian Institute of Technology, Guwahati. 12 Sept. 2013.
2. David Easley and Jon Kleinberg, "Positive and negative relationships," in Networks,
3. Fariss, C. J. & Schnakenberg, K. F. Measuring Mutual Dependence between State Repressive Actions. *Journal of Conflict Resolution* 58**,** 2014.
4. Kim, K. (2018, March 12). *Mathematical approach for Text Mining 1*. Lecture presented in Ulsan National Institute of Science and Technology, Ulsan.
5. Rosario, B. (2000). Latent Semantic Indexing: An Overview (Rep.).
6. Saito, N. MAT 167: Applied Linear Algebra, Lecture 22: Text Mining. 2012.
7. Thedchanamoorthy, G., Piraveenan, M., Kasthuriratna, D., & Senanayake, U. Node assortativity in complex networks: An alternative approach. *Procedia Computer Science, 29*. 2014.
8. Turner, S. Success in Social Movements: Looking at Constitutional-Based Demands to Determine the Potential Success of Social Movements. 2013.

Acknowledgments

This project was mentored by William Lippitt and Heidi Reynolds-Stenson, whose help is acknowledged with great appreciation.