Newman Edge Talk

(Obligatory) Email Statistics

- *The total number of worldwide email accounts is expected to increase from nearly 3.9 billion accounts in 2013 to over **4.9 billion accounts by the end of 2017**. This represents an average annual growth rate of about 6% over the next four years. [Email-Statistics-Report-2013-2017-Executive-Summary]*
- *The total worldwide email traffic, including both Business and Consumer emails, is estimated to be over 215 billion emails/day by year-end 2016, growing to **over 257 billion emails/day by the end of 2020**. [Email-Market-2016-2020-Executive-Summary]*
- *There are approximately 2.5 million emails sent every second [Internet Live Stats]*
- *Spam makes up 61.25 percent of all email traffic worldwide [Statista.com]*

http://www.radicati.com/wp/wp-content/uploads/2013/04/Email-Statistics-Report-2013-2017-Executive-Summary.pdf

(Obligatory) Email Statistics

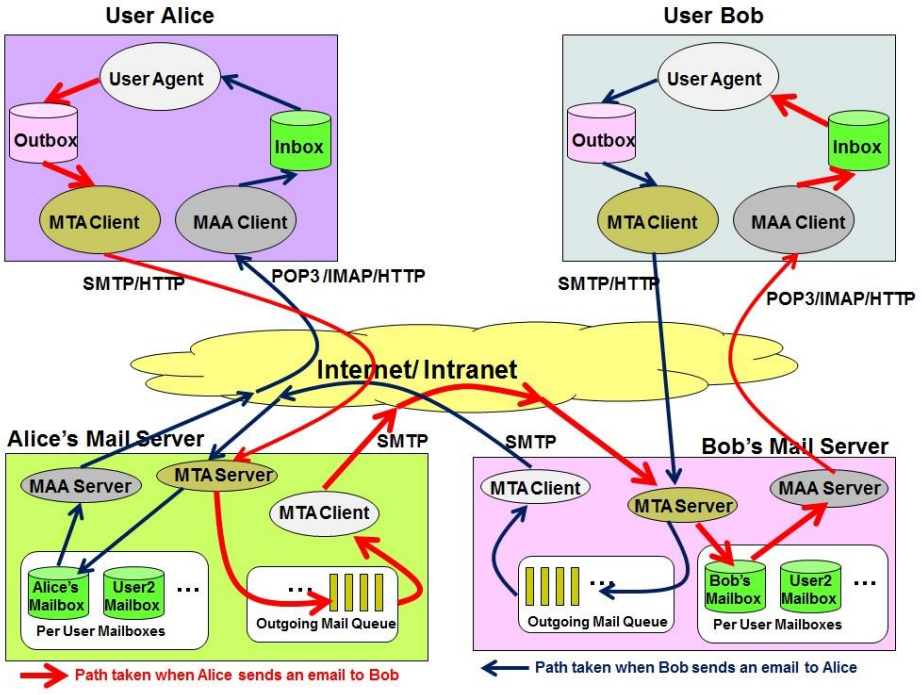
- *In ~2015, the number of business emails sent and received per user per day totals 122 emails per day. This figure continues to show growth and is expected to **average 126 messages sent and received per business user** by the end of 2019 [Email-Statistics-Report-2015-2019-Executive-Summary]*

Business Email	2015	2016	2017	2018	2019
Average Number of Emails Sent/Received per	122	123	124	125	126
Average Number of Emails Received	88	90	92	94	96
Average Number of Legitimate Emails	*76*	*76*	*76*	*76*	*77*
Average Number of Spam Emails	*12*	*14*	*16*	*18*	*19*
Average Number of Emails Sent	34	33	32	31	30

Table 3: Business Emails Sent/Received Per User/Day, 2015 - 2019

Simple Mail Transport Protocol

- SMTP stands for Simple Mail Transport Protocol and operates on port 25. SMTP is a server-to-server protocol. Clients use POP3 or IMAP to retrieve or send messages to the SMTP server, while the SMTP server then communicates to other SMTP servers.
- It is a **text based**, Request – Response, client server protocol, with simple messages like HELO, Mail From, RCTP TO, DATA etc. A sample SMTP session of an email is given in the diagram below



August 1982 RFC 821
Simple Mail Transfer Protocol

APPENDIX A

TCP Transport service

The Transmission Control Protocol [3] is used in the ARPA Internet, and in any network following the US DoD standards for internetwork protocols.

Connection Establishment

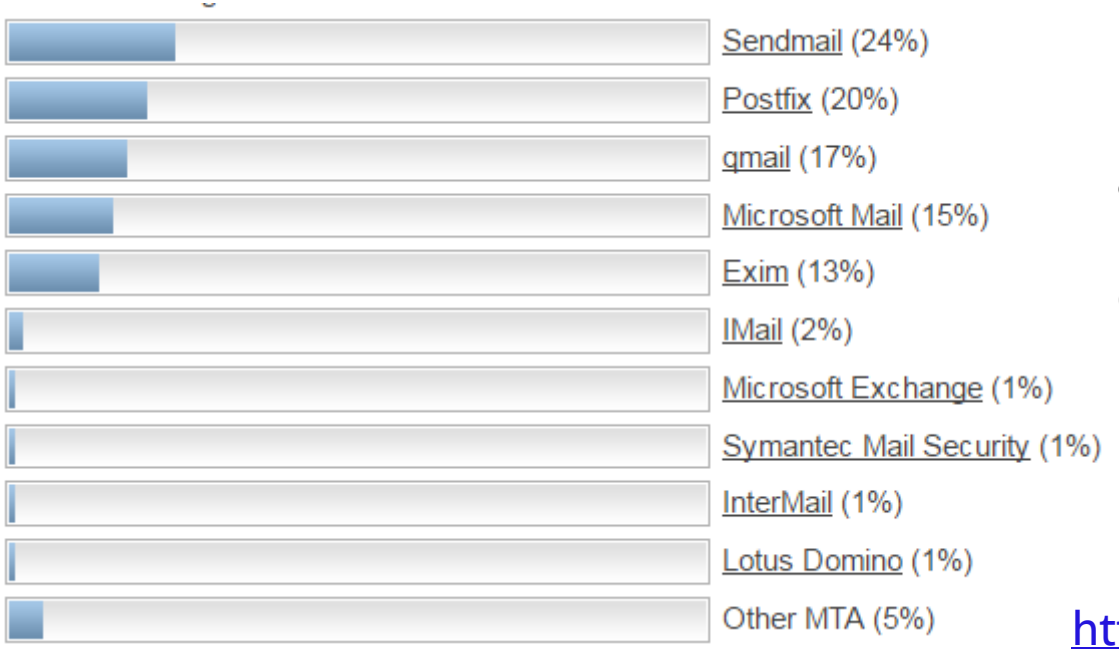
The SMTP transmission channel is a TCP connection established between the sender process port U and the receiver process port L. This single full duplex connection is used as the transmission channel. This protocol is assigned the service port 25 (31 octal), that is L=25.

Data Transfer

The TCP connection supports the transmission of 8-bit bytes. The SMTP data is 7-bit ASCII characters. Each character is transmitted as an 8-bit byte with the high-order bit cleared to zero.

Mail Server

- An Mail Transfer Agent (MTA) implements both the client (sending) and server (receiving) portions of the SMTP
- 2 options:
  - Webmail Service Provider
  - Mail Servers (e.g. MS Exchange, Sendmail)



Exploits exist to determine user accounts, relay email through unsuspecting servers and spoof email user accounts (not MS Ex)

http://www.mailradar.com/mailstat/

Free Email Providers

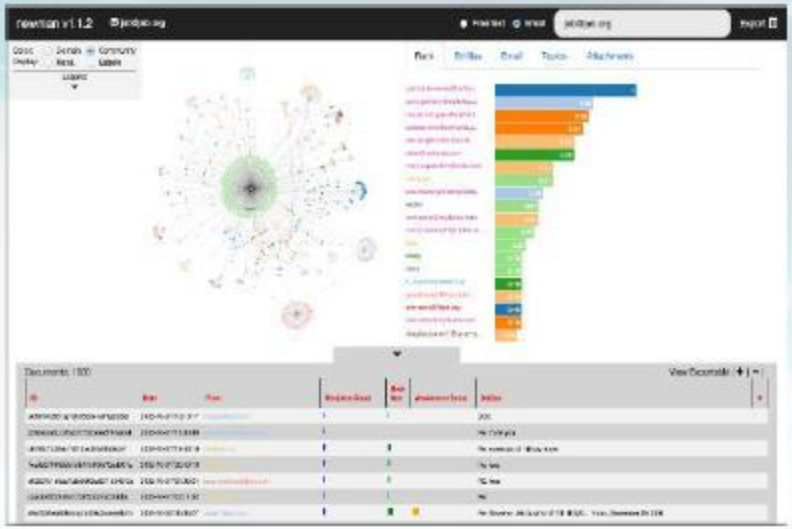
```
123box.net
123india.com
123mail.cl
123qwe.co.uk
150ml.com
15meg4free.com
163.com
1coolplace.com
1freeemail.com
1funplace.com
1internetdrive.com
1mail.net
1me.net
1mum.com
1musicrow.com
1netdrive.com
1nsyncfan.com
1under.com
1webave.com
1webhighway.com
212.com
24horas.com
2911.net
2d2i.com
2die4.com
3000.it
37.com
3ammagazine.com
3email.com
3xl.net
444.net
4email.com
```

Free global-accessible email service providers
http://www.joewein.net/spam/spam-freemailer.htm

Total 2797 including Hotmail, Yahoo, 123.com and those are only those publically listed

Amount of web service providers creates challenges

Newman Tool



Newman is an open source web-based application tool that provides the ability to quickly analyze and explore email using advanced analytics and visualization techniques. These abilities are not possible with traditional email applications. ***Newman*** visualizations help identify trends, patterns, and relationships between entities that may not otherwise be obvious. Visualizations include communication network graphs, email table, email viewer, rank, entity, and topics histograms.

Benefits

Analyze email beyond text search:

- Reduces the amount of time and effort required to analyze an email data set.
- Identify what is pertinent in an email set
- Provide simple and separate visualizations for different facets of the same data, which reduces clutter and information overload.

Technologies

Newman utilizes the following XDATA Open Catalog technologies:

- Tangelo (Kitware)
- MITIE (MIT-LL)
- Topic Clustering (MIT-LL)
- ActiveSearch (CMU)
- Tika (Apache)
- Louvain Modularity (Sotera Defense)

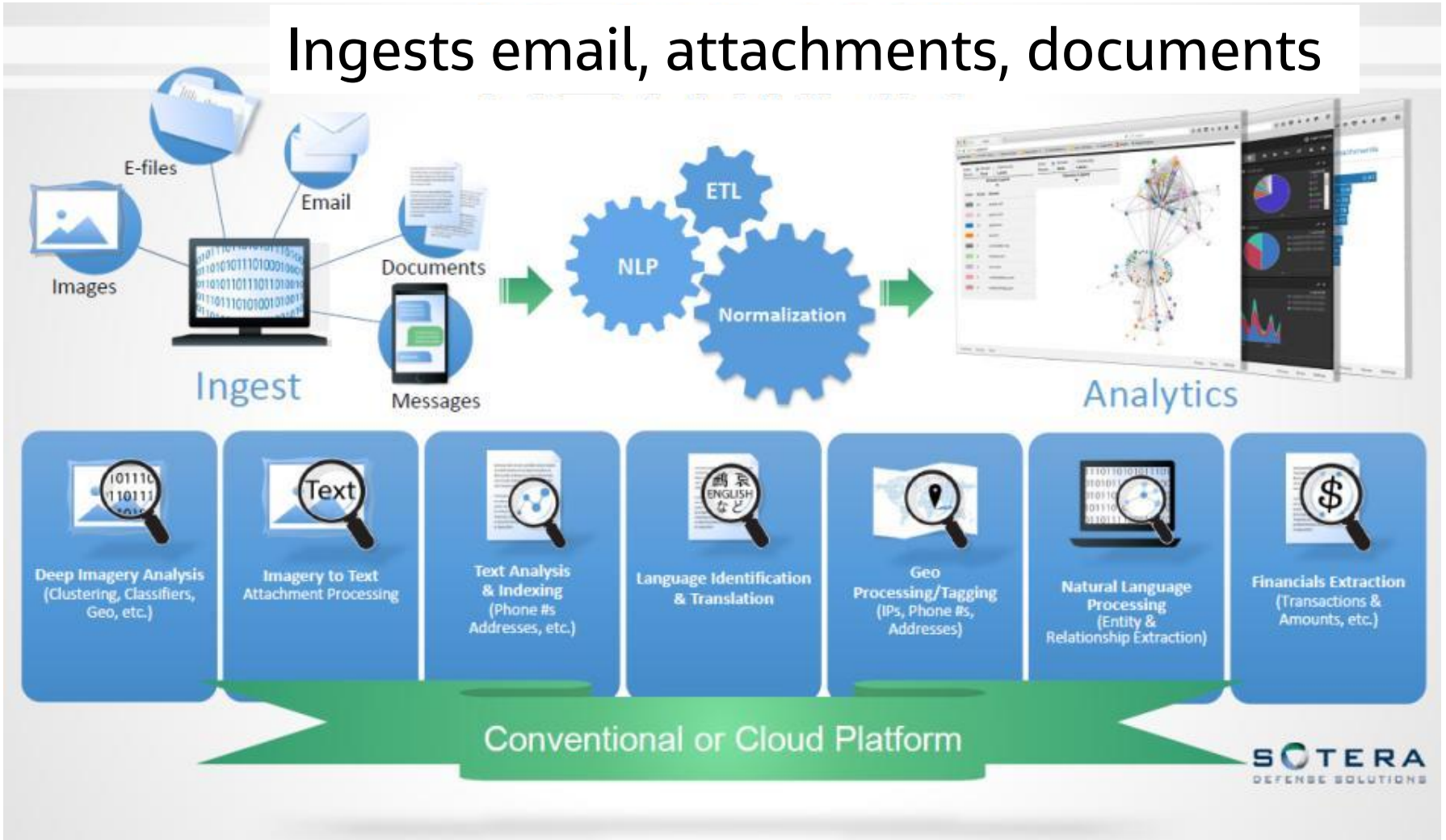
Newman Knows Email

- “Our email spans decades and is too much to analyze. Therefore, we didn’t get approval [to proceed with the email investigation]”
- “What are the most discussed email topics?”
- “Show me all email attachments that contain only pictures of humans in them.”
- “What are the individuals, organizations, locations and phone numbers embedded within discussion?”
- “What communities form over time?”
- “What locations are discussed? Can you display them on a map?”
- “Can you provide a list of email addresses that originate from one location and send to another?”
- “What are the correspondence trends over time between an email community?”

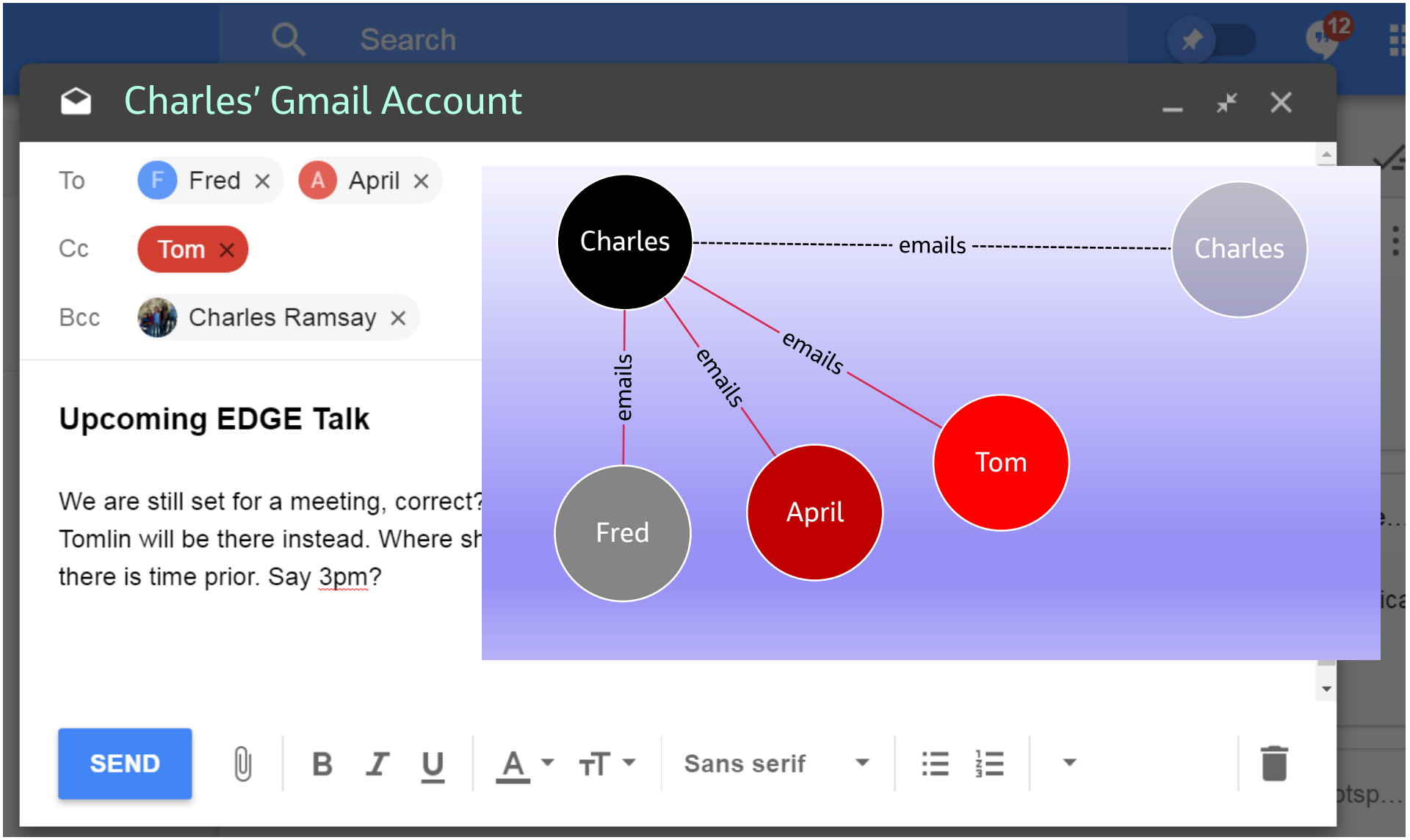
Newman Can…

- Visualize the flow of communication in graphs and maps
- Find sensitive documents and track where they go
- Find the key influencers and those that can connects multiple social networks together
- Provide global search of emails, documents, SMS, images multiple modes of communication
- Search for concepts based on any provided document. Find discussions that relate to a particular theme
- Provide a timeline of all communications between one or more people
- Search over 116 key attributes based on parsed emails, documents, SMS etc

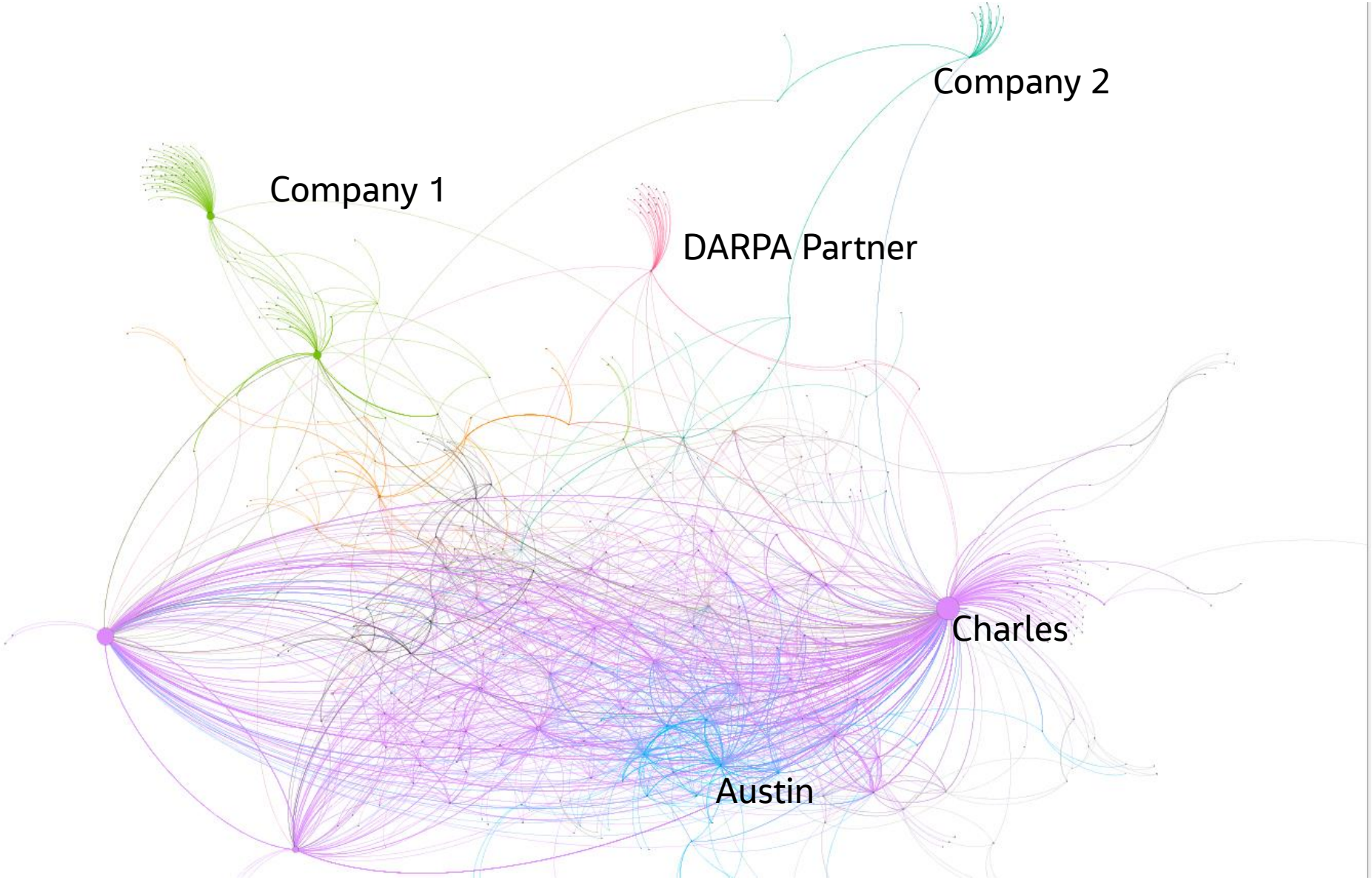
Newman Ingest



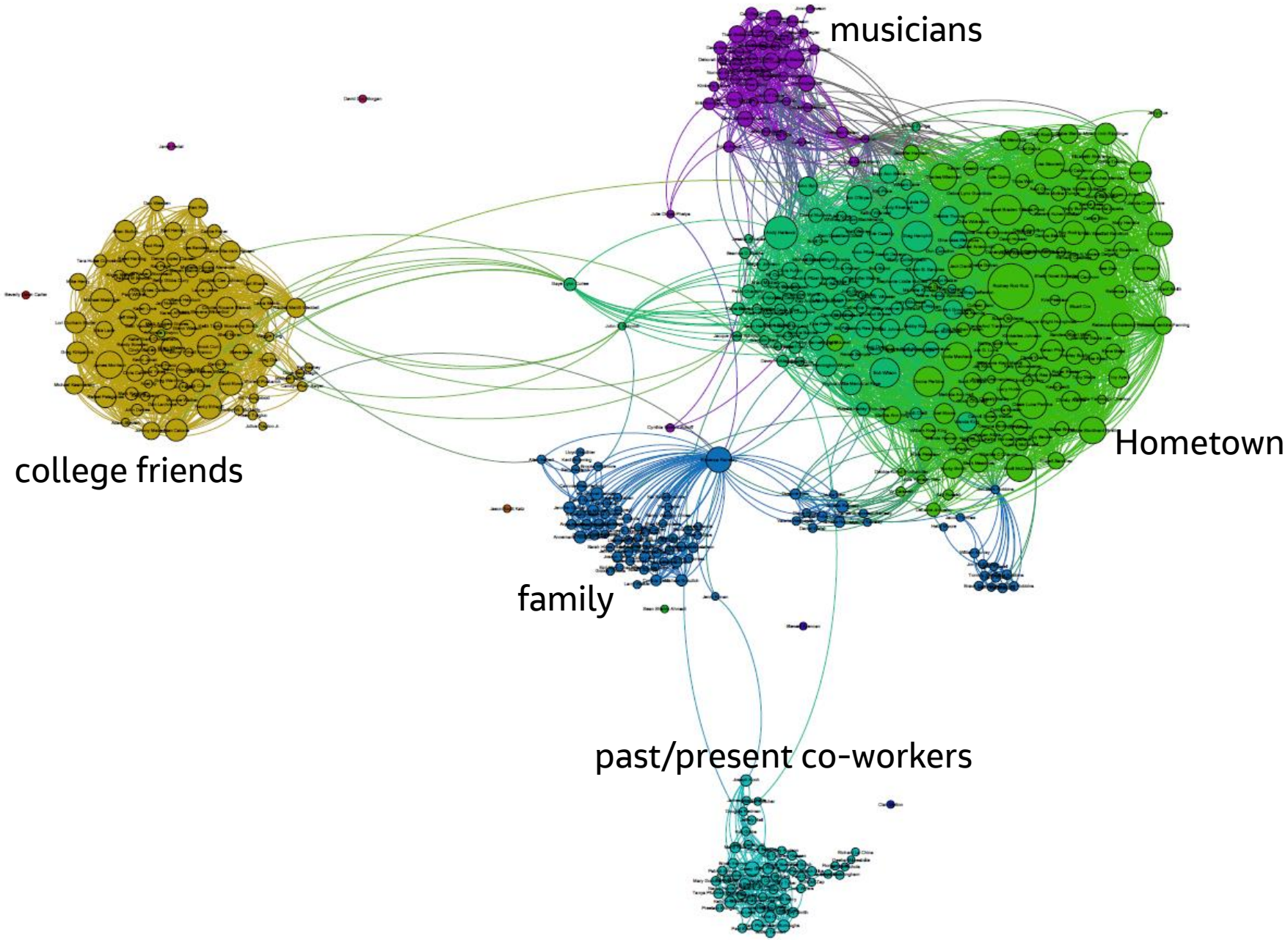
Email Network Representation



Email Network – Charles Communities



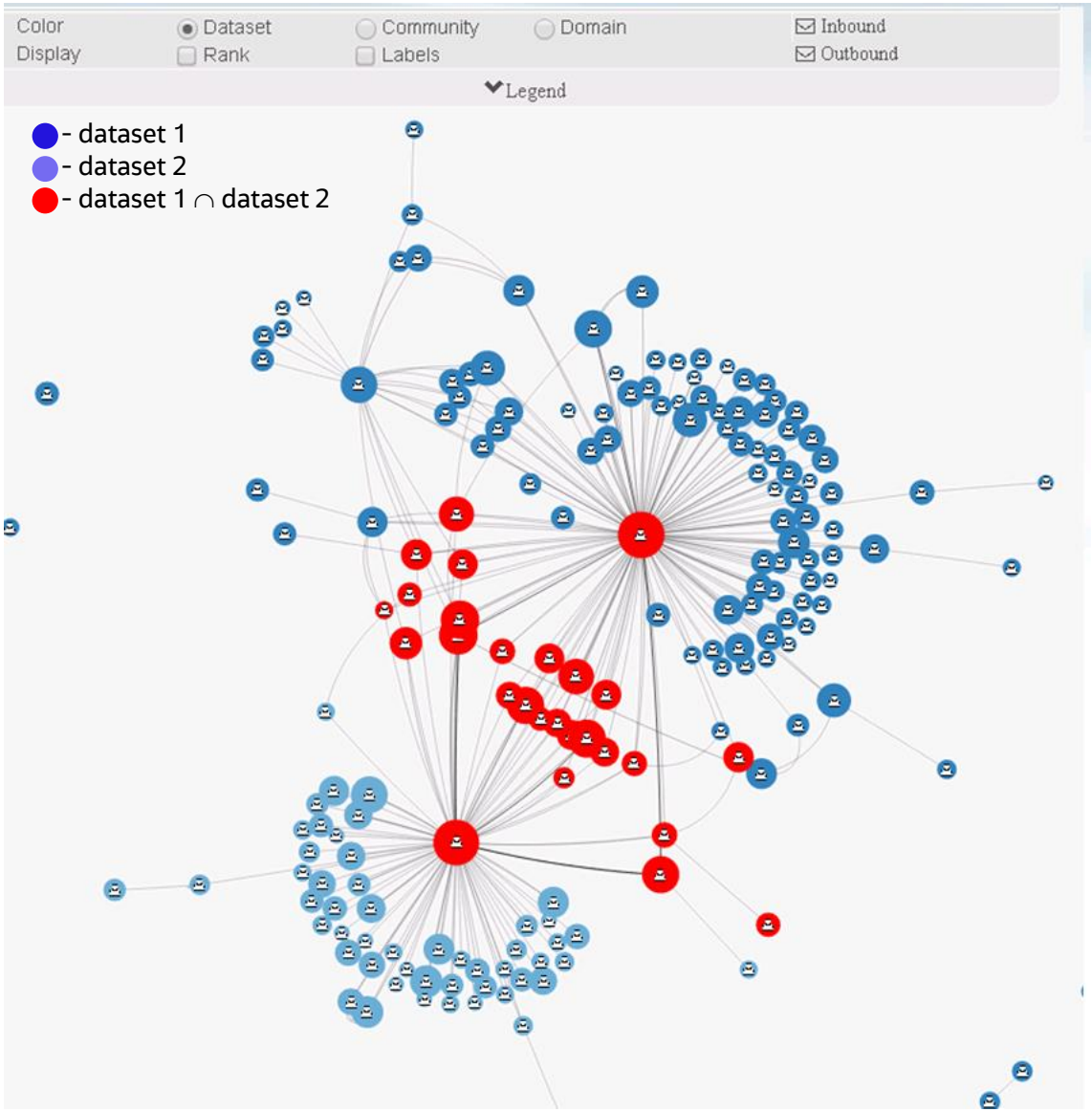
Communities– Facebook Example



Network Graph Analysis

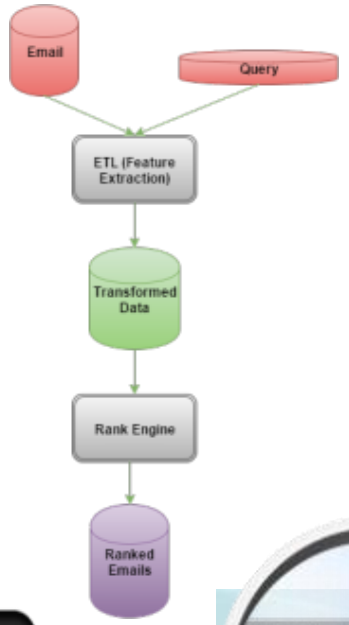
- Newman – Community detection through e-mail communication network graph analysis

- Two distinct communities are identified (light blue, dark blue) and actors common to both communities are identified (red)



List / Document GIST Search Capability

Search using:
- Keywords (e.g. 7 terms)
- Lists (e.g. SDN/CDN)
- Document GIST (Wassenaar)




WA-LIST (15) 1
03-12-2015

THE WASSENAAR ARRANGEMENT

ON

EXPORT CONTROLS FOR CONVENTIONAL ARMS

AND

DUAL-USE GOODS AND TECHNOLOGIES

LIST OF DUAL-USE GOODS AND TECHNOLOGIES

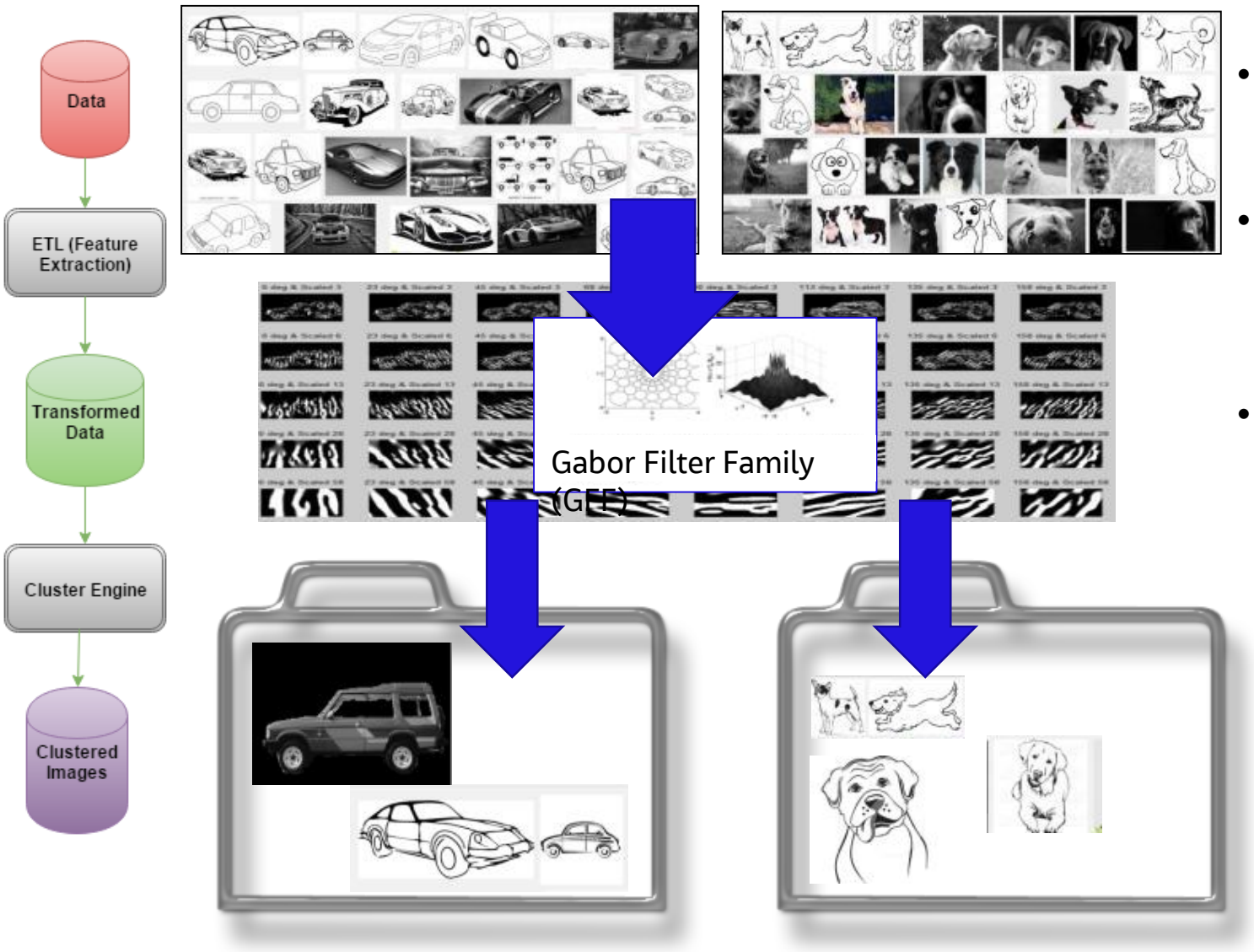
AND

MUNITIONS LIST

(TF-IDF)

Document **"Key" Terms** **Ranked Search Results**

Feature Extraction / Image Clustering



- Mathematically clusters images into groups
- Aids analyst by separating images into natural groups
- Will allow for like-type matches