

CYBERGATE: A DESIGN FRAMEWORK AND SYSTEM FOR TEXT ANALYSIS OF COMPUTER-MEDIATED COMMUNICATION — FIGURES

By: **Ahmed Abbasi**
Sheldon B. Lubar School of Business
University of Wisconsin–Milwaukee
Milwaukee, WI 53217
U.S.A.
abbasi@uwm.edu

Hsinchun Chen
Management Information Systems
Eller College of Management
University of Arizona
Tucson, AZ 85721
U.S.A.
hchen@eller.arizona.edu

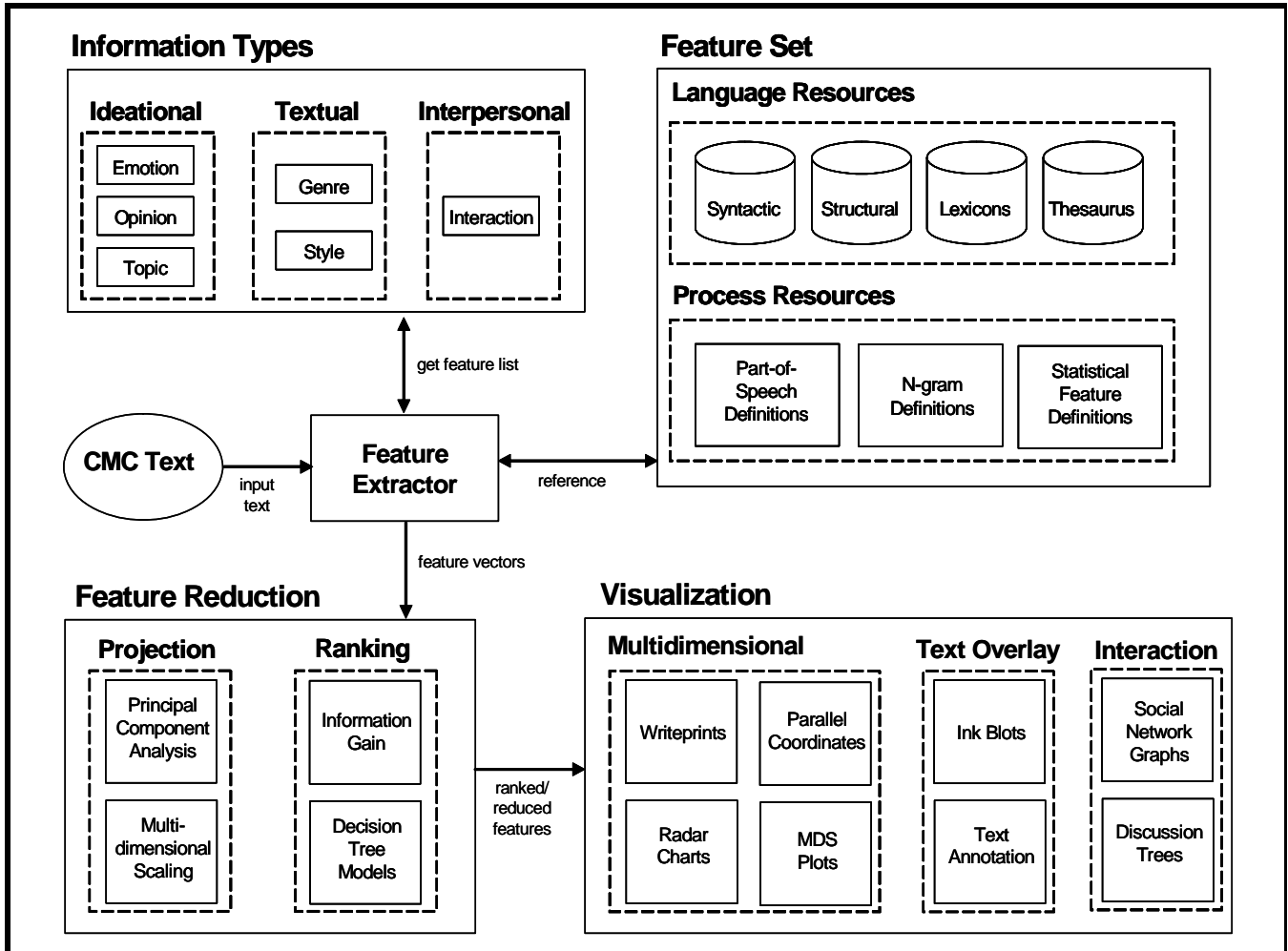


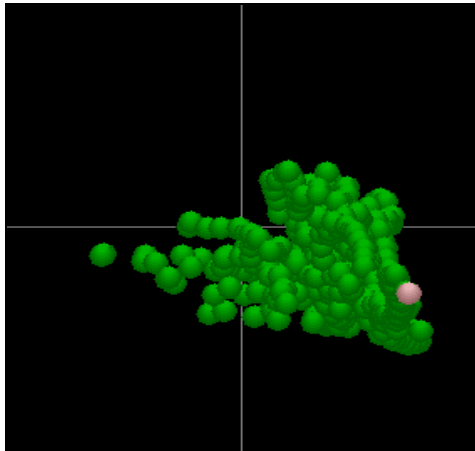
Figure 1. CyberGate System Design

(a) All Features				(b) Projection		(c) Ranking	
No.	Description	Usage	Mean	Ex	Ey	Description	Weight
0	MARK	0.726	0.054	0.278	0.103	PANDA	60.71
1	COUNSELEN...	0.526	0.011	0.099	-0.355	DEVIL	28.85
2	GENERAL	0.526	0.023	0.019	-0.095	LATERAL	18.9
3	PRESIDENT	0.526	0.017	0.227	0.323	KH	14.43
4	TAYLORVICE	0.526	0.011	0.22	0.33	THROUGHPUT	12.14
5	WHOLESALE	0.526	0.041	-0.15	0.06	TW	10.95
6	SERVICES	0.368	0.035	0.152	0.052	SUN	18.9
7	TIME	0.284	0.165	0.247	0.118	GAS	28.85
8	FAX	0.274	0.038	0.078	-0.186	PIPELINE	18.9
9	SMITH	0.274	0.085	0.309	0.013	EVI	21.86
10	TEXAS	0.263	0.095	0.305	0.0040	ECT	5.93
11	STREET	0.253	0.083	0.0080	-0.201	HAT	1.29
12	PARTY	0.221	0.04			ATE	1.0

Figure 2. CyberGate Feature Selection Examples

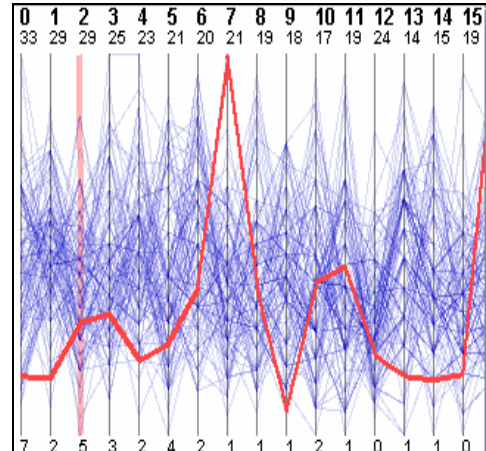
(a) Writeprints

N-dimensional PCA projections based on feature occurrences. Each circle denotes a single message. Selected message is highlighted in pink. Writeprints show feature usage/occurrence variation patterns. Greater variation results in more sporadic patterns.



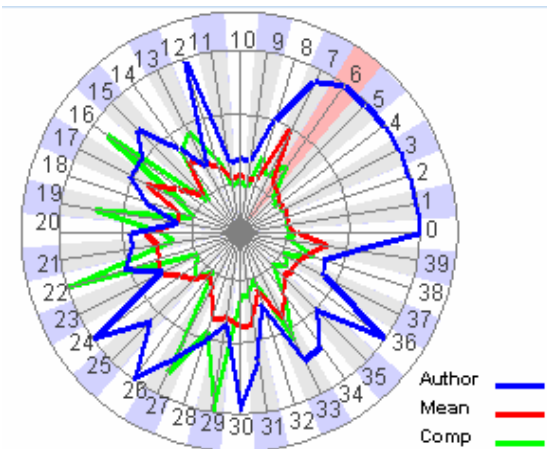
(b) Parallel Coordinates

Parallel vertical lines represent features. Bolded numbers are feature numbers (0-15). Smaller numbers above and below feature lines denote feature range. Blue polygonal lines represent messages. Selected message is highlighted in red. Selected feature is highlighted in pink (#2).



(c) Radar Charts

Chart shows normalized feature usage frequencies. Blue line represents author's average usage, red line indicates mean usage across all authors, and green line is another author (being compared against). The numbers represent feature numbers. Selected feature is highlighted (#6).



(d) MDS Plots

MDS algorithm used to project features into two-dimensional space based on occurrence similarity. Each circle denotes a feature. Closer features have higher co-occurrence. Labels represent feature descriptions. Selected feature is highlighted in pink (the term "services").

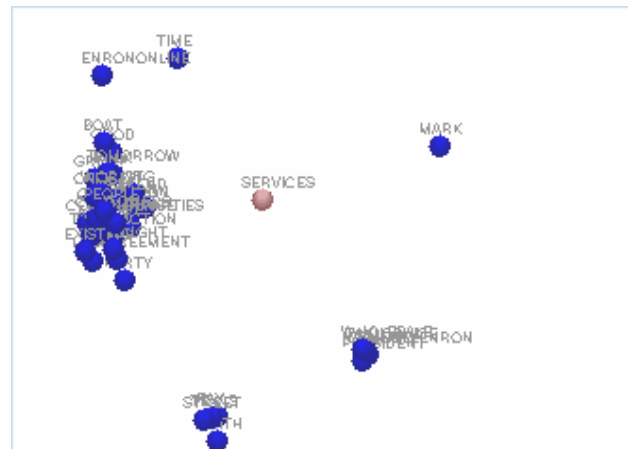


Figure 3. Multidimensional Text Views in CyberGate

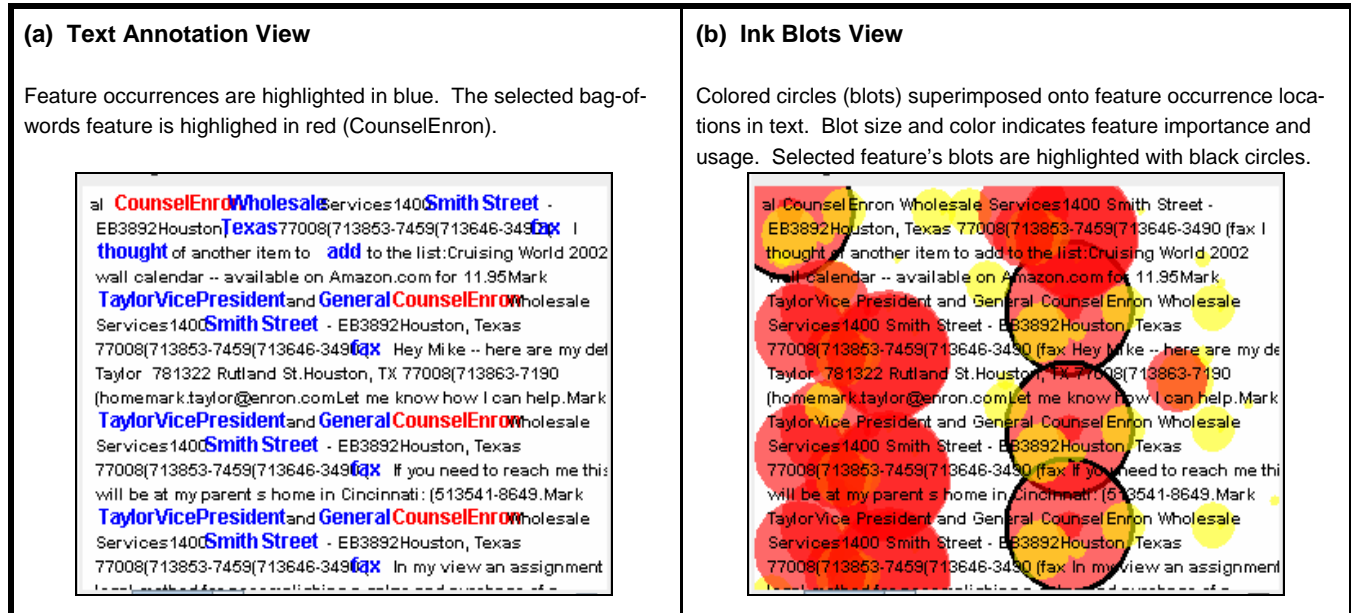


Figure 4. Text Overlay Views in CyberGate

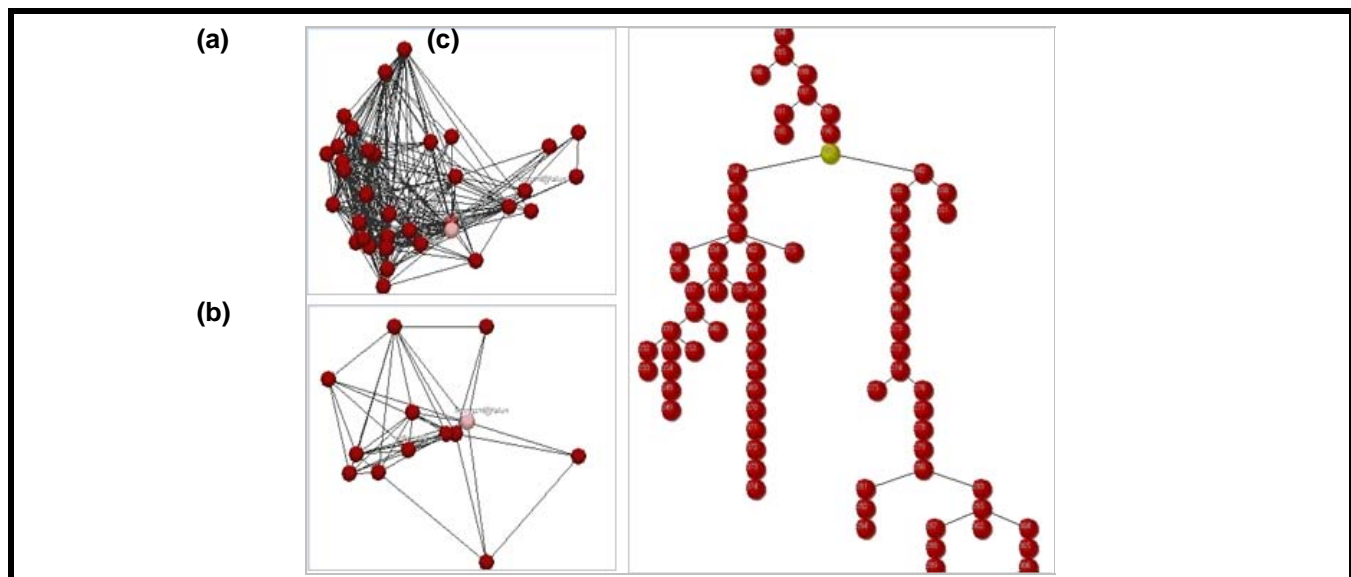


Figure 5. Interaction Views in CyberGate for Representing Interpersonal Information

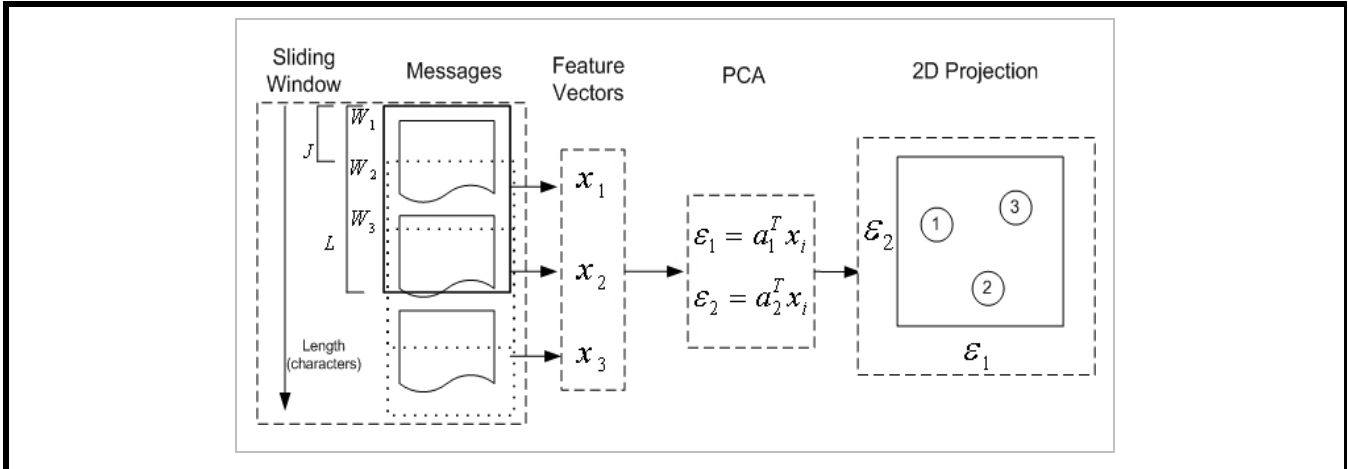


Figure 6. Writenprints Process Illustration on Two Dimensions

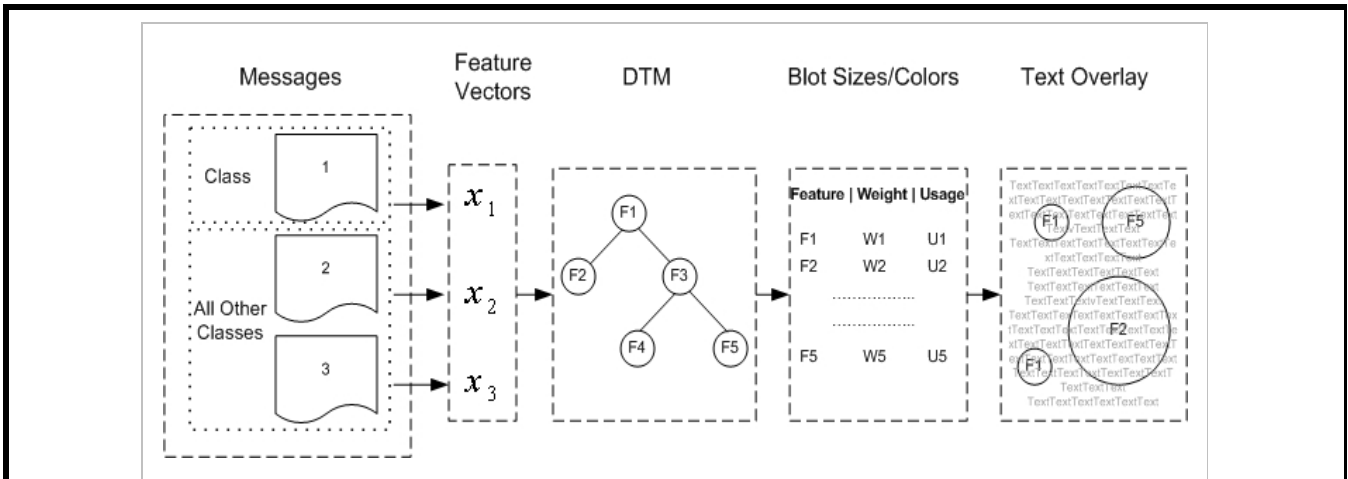


Figure 7. Ink Blots Process Illustration

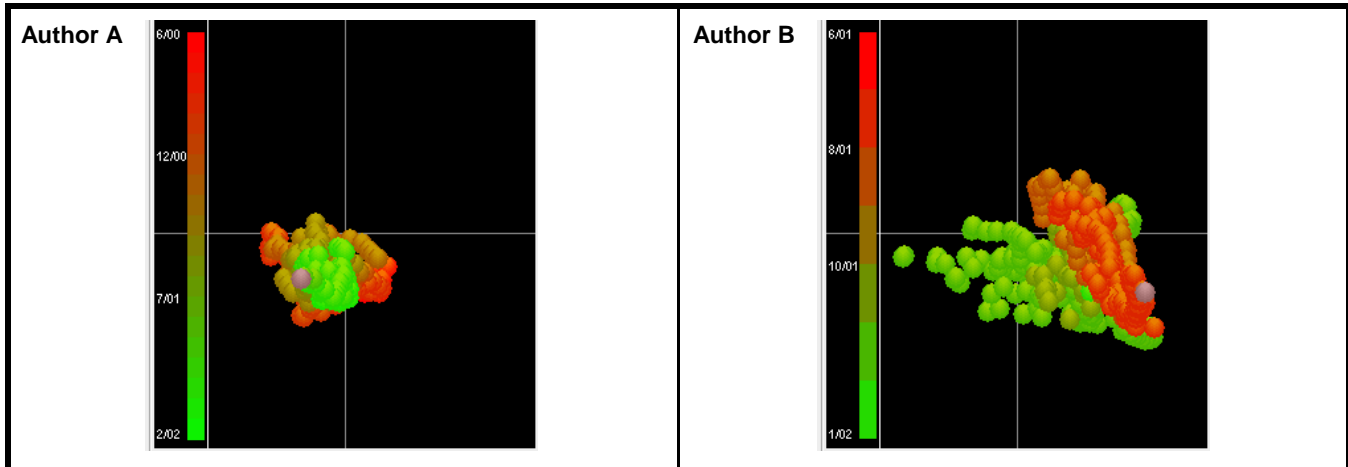


Figure 8. Writeprints for Two Enron Employees

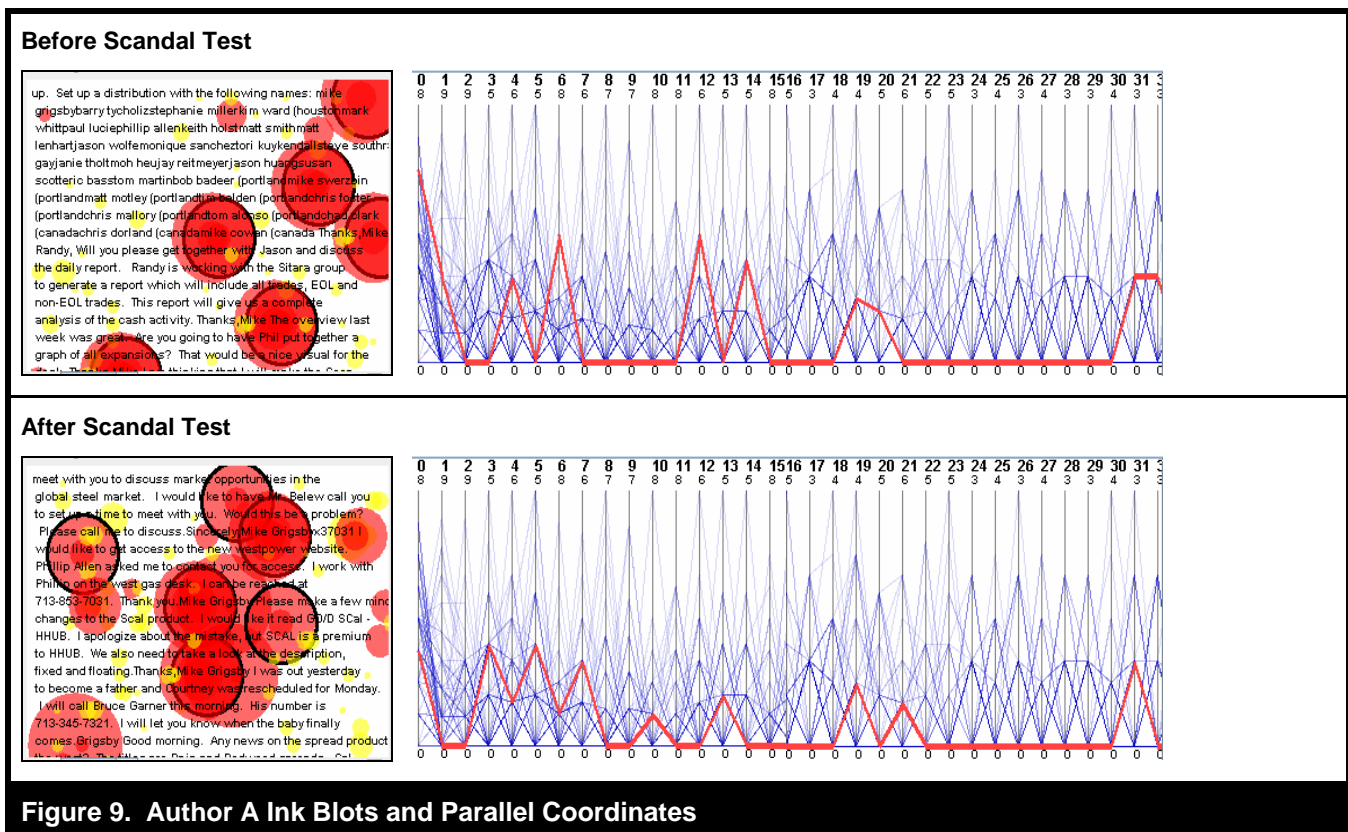


Figure 9. Author A Ink Blots and Parallel Coordinates

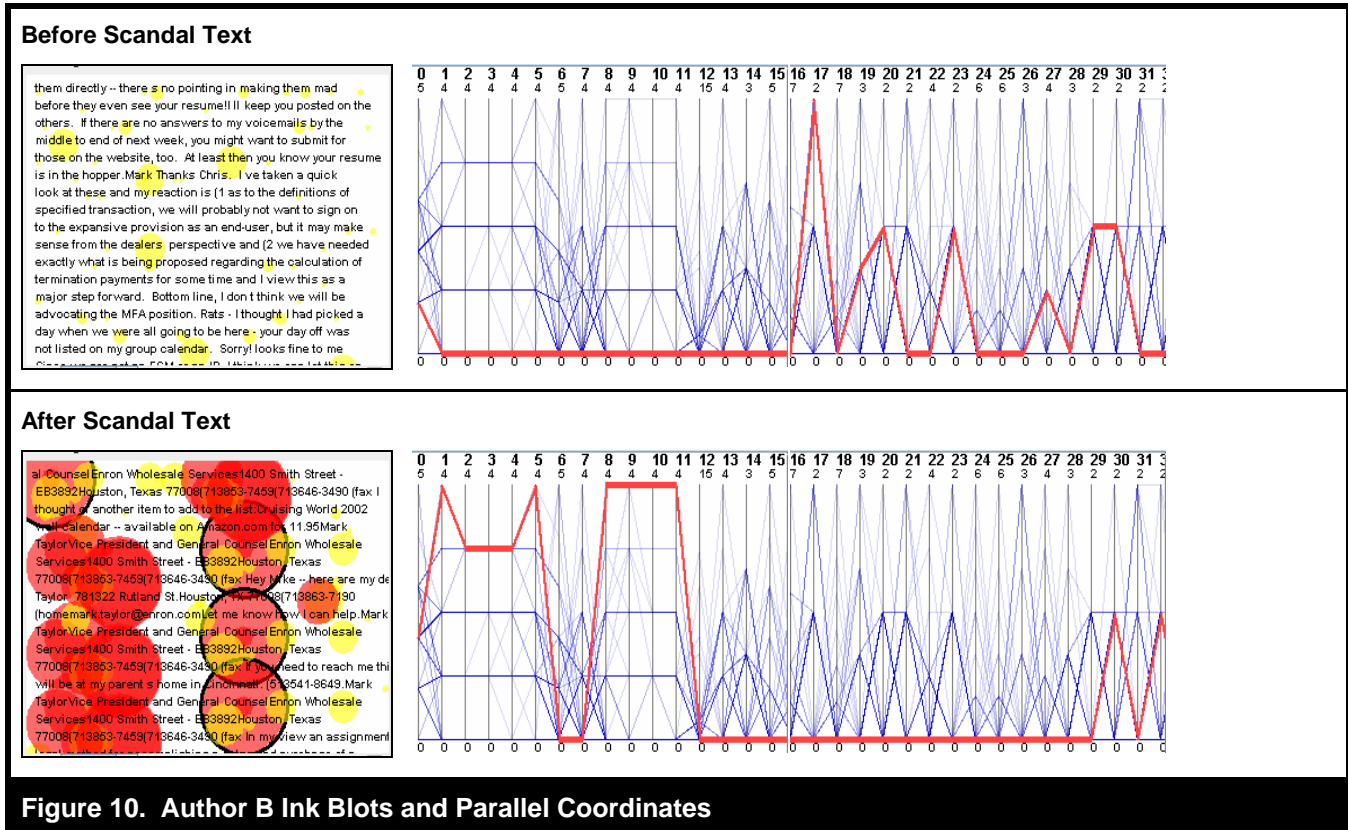


Figure 10. Author B Ink Blots and Parallel Coordinates

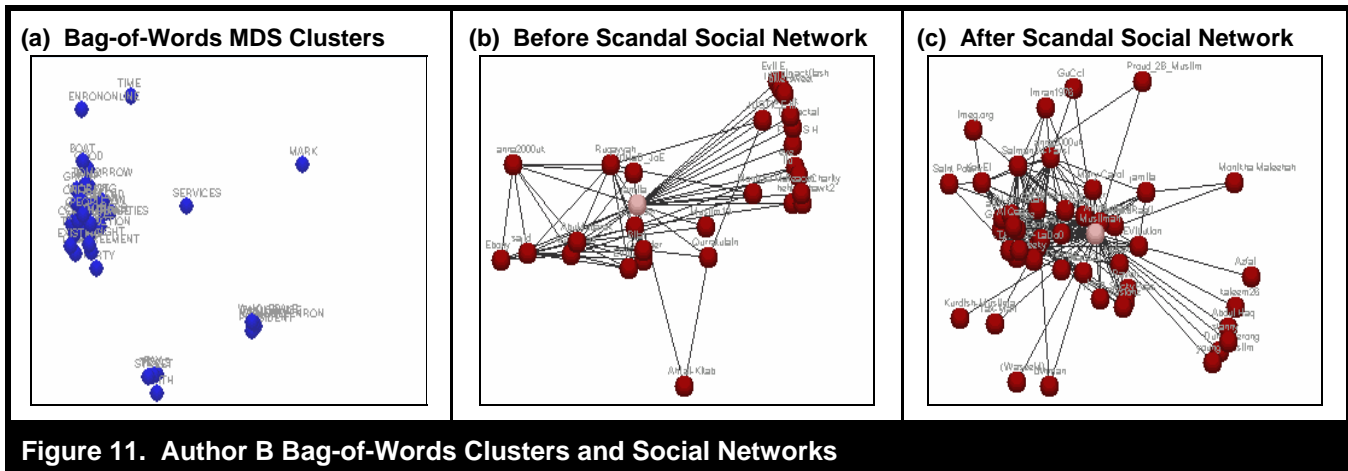


Figure 11. Author B Bag-of-Words Clusters and Social Networks