Editor: Theresa-Marie Rhyne

An Information-Theoretic View of Visual Analytics

Chaomei Chen
Drexel University

Prior to the 9/11 terrorist attacks, several foreign nationals enrolled in US civilian flying schools to learn how to fly large commercial aircraft. They wanted to learn how to navigate civilian airlines, but they were not interested in landings or takeoffs. They all paid cash for the lessons. So, the 9/11 investigations raised questions about whether intelligence agencies could have connected the dots and prevented the attacks.[1] But how do you connect these seemingly isolated dots and reveal the hidden story?

One clue might lie in the differentiation between puzzles and mysteries that Malcolm Gladwell made in a recent *New Yorker* article on stories about Enron's collapse.[2] To solve a puzzle, Gladwell writes, you need a specific piece of information; but to solve a mystery, you must ask the right question. Connecting dots is more a mystery than a puzzle. You might have all the necessary information in front of you and yet fail to see the connection or recognize an emergent pattern. Asking the right question is critical to staying on track.

Visual analytics is an emerging discipline that helps connect dots. It facilitates analytical reasoning and decision making through integrated and highly interactive visualization of complex and dynamic data and situations.[3] Solving mysteries is only part of the game. Visual analytics must augment analyst and decision-maker capabilities to assimilate complex situations and reach informed decisions. Information theory offers a framework for keeping focused on the right questions.

An information-theoretic view

In information theory, the information value carried by a message is the difference in information entropy before and after receipt of the message. Information entropy is a macroscopic measure of uncertainty defined on a frequency or probability distribution. The information-theoretical approach attempts to quantify discrepancies of the information content of distributions. Information indices, such as the widely known Kullback–Leibler divergence,[4] are entropy-based measures of discrepancies between distributions.[5]

We can define the K-L divergence of probability distribution Q from probability distribution P as

$$\text{Divergence}_{K-L}(P:Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

The divergence measures the loss of information if we use Q instead of P, assuming P is the true distribution. We can view information entropy as the divergence from a uniform distribution. This is consistent with the common interpretation of information entropy as a measure of uncertainty, or lack of uniformity.

A useful alternative interpretation of the K-L divergence is the expected extra message length the communication entails if the message is transmitted without using a coding scheme based on the true distribution. In computer science, an object's *Kolmogorov complexity*, also known as algorithmic entropy or program-size complexity, measures the amount of computational resources needed to specify the object.

Visual analytics must integrate and interpret findings at both microscopic and macroscopic levels. Human analysts are good at dealing with macroscopic patterns and mysteries, but they can be overwhelmed by microscopic attributes and relations. Information-theoretic strategies and techniques can help. They lend themselves to many ways of decomposing and aggregating information across microscopic and macroscopic levels of abstraction—for example, statistical decomposition analysis. Information-theoretic analysis provides a unifying foundation for studying uncertainty, belief, and evidence in complex situations.

Information foraging and sense making

Situation awareness is a critical part of analytical reasoning. It typically involves two integral and iterative subprocesses, information foraging and sense making, which we can characterize from an information theoretical point of view.

Information-foraging theory is a predictive model for search behavior.[6] According to the theory, an information environment is made of patches of information; an information forager moves from one patch to another in search of information, just as a predator searches for its prey. The theory answers questions about how foragers choose information patches and how they spend their time with them.

Information-foraging theory assumes that people adapt search strategies to maximize their profitability, or profit-to-investment ratio. The *investment* typically includes the time spent in searching and assimilating information from patches. The *profit* includes the gain in finding relevant information. Users, or information foragers, tend to follow a path that can maximize overall profitability. *Information scent* is the perception of the value, cost, or accessible path of information sources. It

 Published by the IEEE Computer Society

gives users clues on their next move. When possible, we rely on it to estimate the potential profitability of an information patch.

The power of information-foraging theory is its adaptability and extensibility. It provides a quantitative framework for interpreting behavioral patterns at both microscopic and macroscopic levels. For instance, connecting the dots of mysterious behaviors of 9/11 hijackers at flying schools would depend on the relevant scent's prevalence and strength.[1] The next question to answer is where and how an analyst could find the right information scent to follow in the first place.

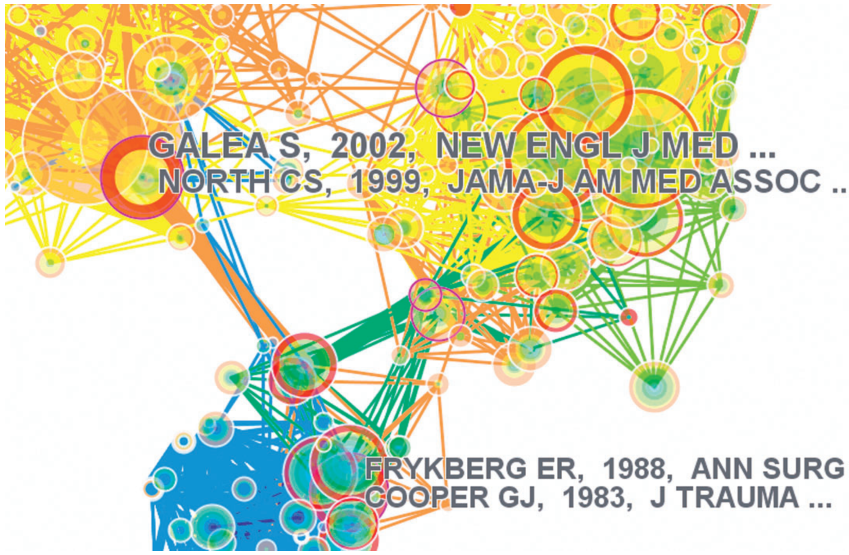


Figure 1. Three clusters of co-cited papers on terrorism research. Each circle depicts a paper. Lines connecting papers are co-citation links. Prominently labeled papers in each patch offer the patch's topical scent. Patch colors, indicating the time of a connection, provide a scent of freshness. Citation ring sizes provide a scent of citation popularity.

Figure 1 is not designed with information-foraging theory in mind, but the visualization intuitively illustrates the profit maximization principle behind the theory. The connective density reinforces the patch boundaries. Colors and shapes give various information scents about each patch. These scents will help users choose which patch they want to explore in more detail.

From an information-theoretic view, information scent only makes sense if it's connected to the broader information-foraging context, including the search's goal, the analyst's or information forager's prior knowledge, and the contextual situation. This broader context implies a deeper connection between the information-theoretic view and the various analytic tasks in situation awareness.

For example, voting in political elections involves a complex sense-making and reasoning process. Voters must make sense of overwhelmingly diverse information, differentiate political positions, accommodate conflicting views, adapt beliefs in light of new evidence, and make macroscopic decisions. Information-theoretic approaches provide a valuable, generic strategy for addressing these issues. Voters are influenced by candidates' positions regarding a spectrum of political issues and their own interpretations of those positions.[7]

Researchers are particularly interested in how uncertainties about candidates' political positions affect voters' decisions. Information theory lets us represent candidate positions on controversial issues as a probability distribution. The underlying true distribution is unknown. We can measure the voters' uncertainty by the divergence of a sample from the true distribution. In a 1980 US presidential-election study, Larry Bartels found that voters in general dislike uncertainty.[8]

In a 1994 congressional-election study, Jeff Gill constructed an aggregate measure of uncertainty about candidates as well as political issues based on three levels of voter certainty regarding each political issue.[7] The study analyzed answers from 1,795 respondents on issues such as crime, government spending, and healthcare. The results suggested that politicians are better off with unambiguous positions, provided their positions do not drastically differ from those held by widely supported candidates.

The effect of uncertainty seems to act at aggregated levels as well as individual levels. Crime, for example, has been a Republican campaign issue for decades. Gill's study found that Republican candidates who are vague on this issue almost certainly suffer at the polls.[7] This example shows that an information-theoretical approach provides a flexible tool for studying information uncertainty in complex reasoning and decision-making processes.

Evidence and beliefs

People review their beliefs when new information becomes available. For example, physicians run various tests on their patients, then they make sense of test results and decide if more tests are needed. In general elections, voters ask questions about candidates' political positions to reduce or eliminate uncertainties about choosing among them. *Bayesian reasoning* is a widely used method to analyze evidence and synthesize our beliefs. Applications cover a wide variety of domains, from interpreting mammography for breast cancer risks to differentiating spam from genuine email. US Navy investigators successfully applied Bayesian reasoning to the search for the *USS Scorpion*, a nuclear submarine that was lost at sea in May 1968. The search was particularly challenging because no one knew its location prior to its disappearance. The Bayesian search theorem guided the search according to the following six steps:

1. Formulate hypotheses of whereabouts of a lost object.
2. Construct a probability distribution over a grid of areas based on the hypotheses.
3. Construct a probability distribution of finding the lost object at a location.
4. Combine the two distributions and form a probability distribution; use the new distribution to guide the search.
5. Start the search from the area with the highest prob-

ability and move to areas with the next highest probabilities.
6. Revise the probability distribution using the Bayesian theorem as the search goes on.

In the *USS Scorpion* search, experienced submarine commanders independently provided hypotheses of the *Scorpion*'s whereabouts. Investigators started the search from the grid square of the sea with the highest probability of finding the submarine and moved on to squares with the next highest probabilities. They used the Bayesian theorem to upgrade the probability distribution over the grid as the search moved along. They found the *Scorpion* in October 1968 more than 10,000 feet underwater, within 200 feet of the location suggested by the Bayesian search theorem.

Bayesian reasoning lets searchers estimate a search's cost at local levels and adapt their search path according to revised beliefs as the process progresses. This adaptive strategy is similar to the profit maximization assumption of information-foraging theory. The revision of beliefs turns probabilistic distributions to information scents. The Bayesian search method is a tool that might help analysts solve mysteries.

Saliency and novelty

Solving mysteries in visual analytics is akin to finding needles in haystacks. Although success finally depends on human analysts to identify subtle information differences, the analysts need tools that can reliably single out subtle outliers or surprises from an overwhelmingly vast and diverse population. *Information indices* capture feature discrepancies in different distributions.

Saliency and novelty are essential information properties to visual analytics. A salient feature or pattern is prominent in that it stands out perceptually, conceptually, or semantically. In contrast, novelty characterizes the uniqueness of information. A landmark has a high saliency in its skyline, whereas a design's novelty measures how unique it is in comparison to others. Humans can often effortlessly spot salient or novel features visually. However, identifying these features computationally is difficult because they are emergent and macroscopic as opposed to specific and microscopic. Capturing emergent features and matching semantic features with visual saliency and novelty is a fundamental challenge in visual analytics.

Information theory can be used to define saliency as statistical outliers in a semantic or visual feature space. Novelty, on the other hand, can be defined as statistical outliers along a specific dimension of the space, such as the temporal dimension.

Using saliency and novelty, Laurent Itti and Pierre Baldi developed a computational model that can detect surprising events in video.[9] They define surprising scenes in terms of the piece of data that is responsible for how an individual changes a belief between two distinct frames. A belief is transformed from a prior distribution $P(\text{Model})$ to a posterior distribution $P(\text{Model} \mid \text{Data})$. They measure the difference between prior and posterior distributions over all models by relative entropy, that is, the K-L divergence. Surprise is then the average of the log-odd ratio with respect to the prior distribution over the set of models. The higher the K-L scores, the more discriminate the detection measures are. Surprises identified by the computational model turned out to be a good match to human viewers' eye movements on video images.

Structural holes and brokerage

Surprises are surprises because they are not expected in a particular context. Similarly, there is an interesting connection between context and creativity. An idea that one community sees as trivial or common might seem inspirational and creative in another. Where do we expect to find creative ideas in our society, an environment full of information patches?

Structural holes are a topological property of a social network. According to Ronald S. Burt, structural holes refer to the lack of comprehensive connectivity among social-network components.[10] Person-to-person links are few at structural holes, whereas the links tend to be uniformly strong at a group's center. Because structural holes restrict information flows, they offer potential advantages to the privileged few who are strategically positioned over them.

The advantages are called a vision advantage. People connected across groups are more familiar with alternative ways of thinking, which gives them more options to select from and synthesize. Because they have more ideas to choose from, the selection quality tends to be better.

Burt identifies four brokerage levels for creating value in this situation, from the simplest to the most advanced:

1. Increasing the mutual awareness of interests and difficulties of people on both sides of a structural hole.
2. Transferring best practice between two groups.
3. Drawing analogies between groups seemingly irrelevant to one another.
4. Synthesizing thinking and practices from both groups.

Burt found indeed that a vision advantage associated with brokerage translates to better-received ideas.

A brokerage between information foraging and the structural holes theory might be fruitful for visual analytics. The structural-hole-induced brokerage perspective addresses situations where information scents are either missing or unreachable. On the other hand, the structural-hole theory can guide foragers in selecting potentially information-rich paths. Consider the three prominent clusters shown in Figure 1—a brokerage perspective focuses on the linkages connecting distinct clusters. Consequently, the focus on intercluster connections provides unique leverage to differentiate individual papers at a higher aggregation level. In the terrorism research example, the earliest theme is about physical injuries; a later theme is centered on healthcare and emergency responses; the most recent theme focuses on psychological and psychiatric disorders. Cognitive transitions from one theme to another become easier to grasp at this level. This high-level understanding can also serve as a meaningful context for detecting what is

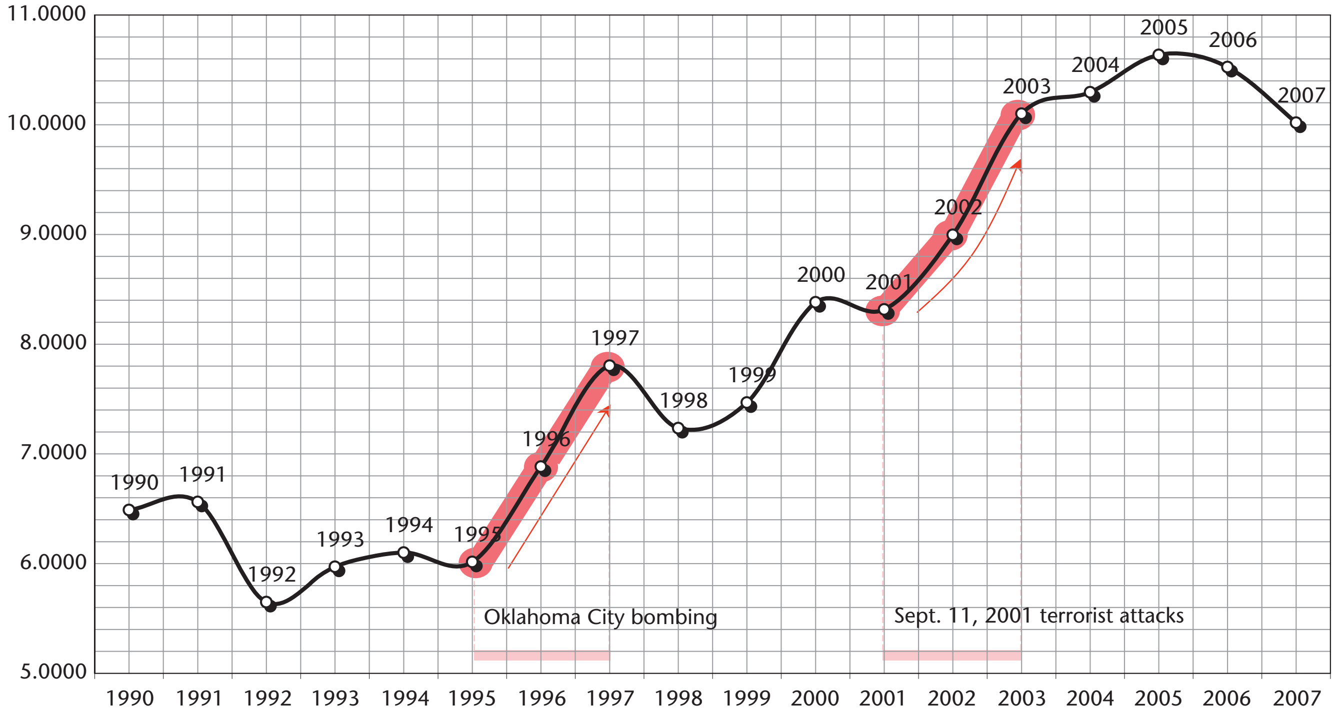


Figure 2. Information entropies of the terrorism-research literature between 1990 and the first half of 2007. The two steep increases correspond to the Oklahoma City bombing of 1995 and the 9/11 events.

common and what is surprising. To understand terrorism research as a whole, we must understand how these themes interrelate. The whole here is more than the sum of the parts. An information-theoretic view brings us a macroscopic level of insights.

Macroscopic views of information content

Information entropy is a useful system-level metric for information uncertainty in a large-scale dynamic system. Figure 2 shows entropy fluctuations in terrorism research on the basis of keywords assigned to scientific papers between 1990 and the first half of 2007. Entropies are computed retrospectively according to the accumulated vocabulary throughout the entire period. The graph shows two steep increases in 1995–97 and 2001–03. These increases indicate fundamental changes in the overall landscape of terrorism research. Information-theoretic insight uniquely identifies emergent macroscopic properties without overwhelming analysts with much microscopic detail. Using the terminology of information foraging, these two periods have transmitted the strongest information scent. Note that using the numbers of unique keywords fails to detect the first period identified by information entropy. Subsequent analysis at microscopic levels reveals that the two periods are associated with the Oklahoma City bombing in 1995 and the 9/11 events.

Information indices let us compare the similarity between different years. Figure 3 (next page) shows a 3D surface of the K-L divergences between distributions in different years. The higher the elevation, the more difference there is between two years of research. For example, the blue area has the lowest elevation, which means that research is more similar in the recent three years than in earlier years.

Information-theoretic techniques provide not only a means of addressing macroscopic questions but also a way to decompose and analyze questions at lower aggregation levels. Given that we have seen two periods of fundamental transformation in terrorism research, the next step is to understand what these changes are in terms of their saliency and novelty. Different distributions might lead to the same entropy level. To compare and differentiate distributions, we can use information-theoretic metrics such as *information bias*, which measures the degree to which a subsample differs from the entire sample that it belongs to. We can easily identify high-profile thematic patterns in terms of term frequencies. Low-profile thematic patterns are information-theoretic outliers from the mainstream keyword distributions. Low-profile patterns are as important as high-profile patterns in analytical reasoning because they tell us something that we are not familiar with, and so something novel.

Informational bias $T(a{:}B)$ is defined as follows, where a is a subsample of the entire sample B, and b is a member of B. p_{ab}, p_a, p_b, and $p_{b/a}$ are corresponding probabilities and conditional probabilities. $H_a(B)$ is the conditional entropy of B in the subsample. We can compare the temporal distribution of a given keyword to the entire space of temporal distributions of keywords:

$$T(a:B) = \frac{1}{p_a}\sum_b p_{ab}\log_2\frac{p_{ab}}{p_a p_b}$$

$$T(a:B) = \sum_b p_{b|a}\log_2\frac{p_{b|a}}{p_b} = -\sum_b p_{b|a}\log_2 p_b - H_a(B)$$

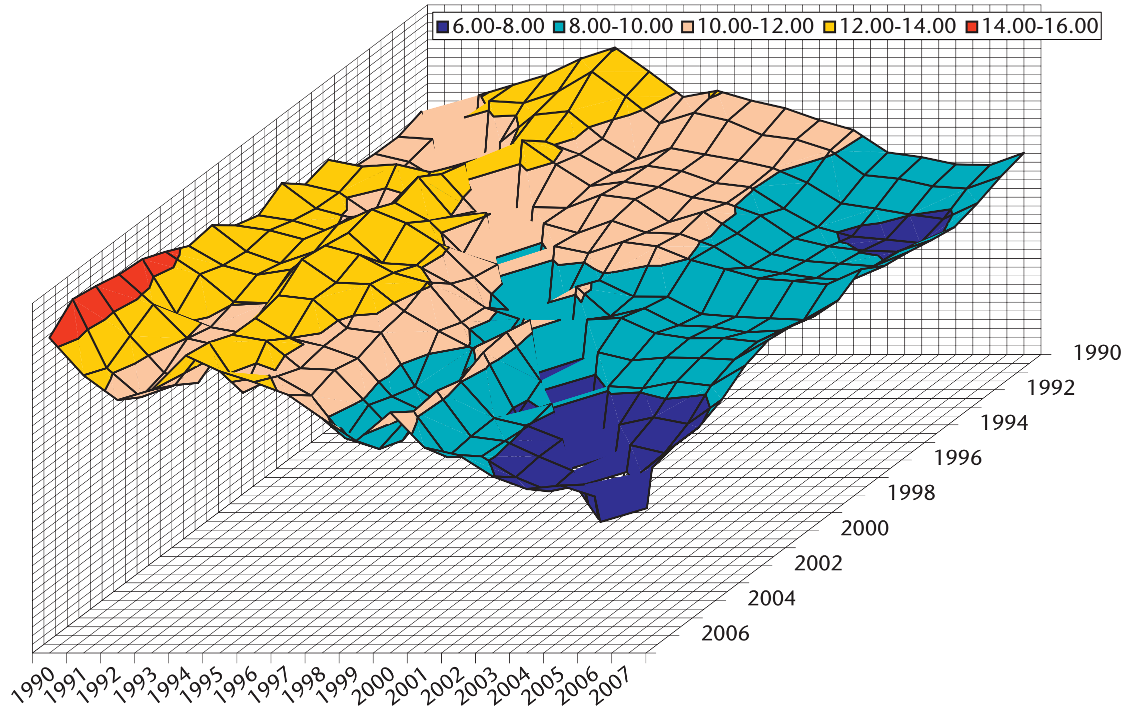
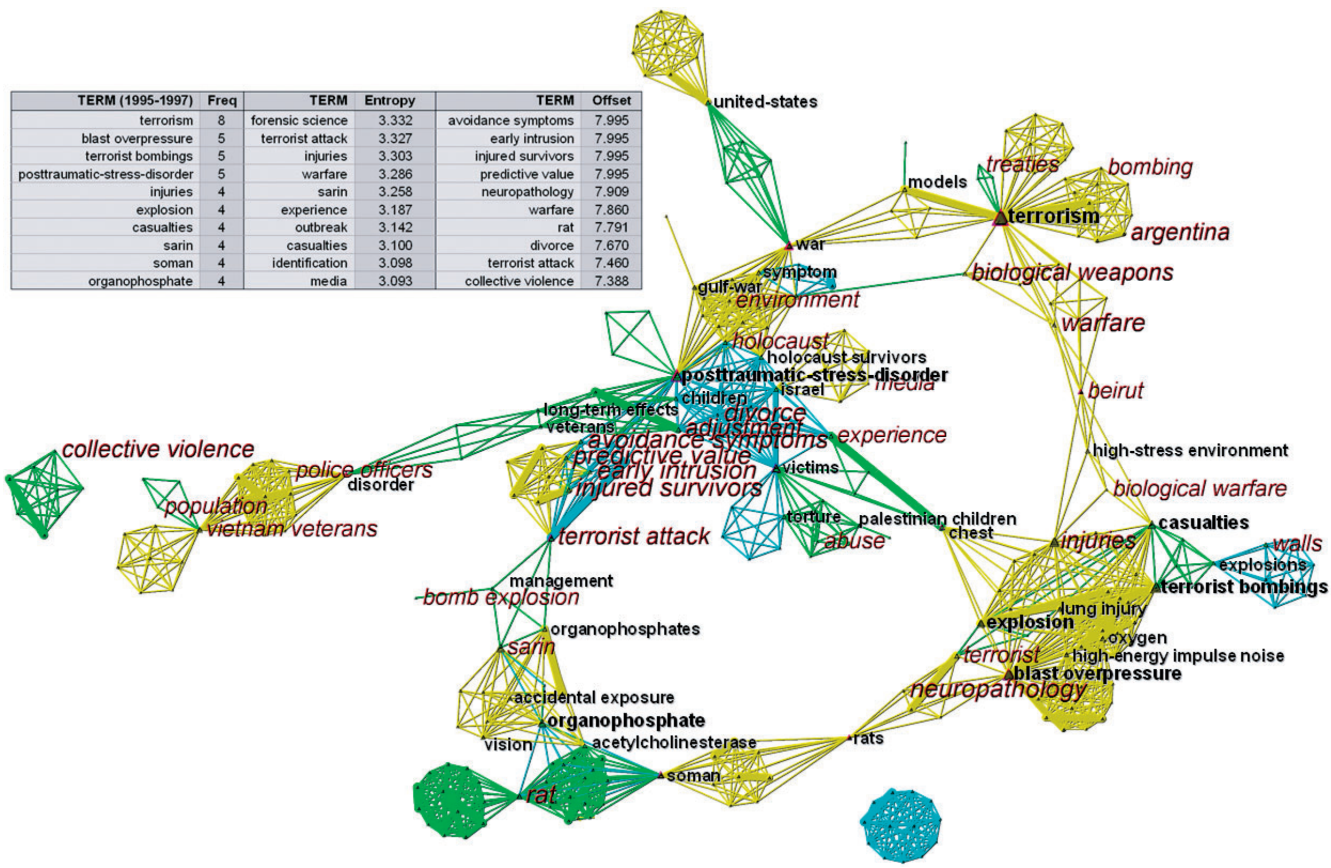


Figure 3. Symmetric relative entropy matrix shows the divergence between the overall use of terms in the terrorism-research literature across different years. The recent few years are most similar to each other. The boundaries between areas in different colors indicate significant changes of underlying topics.



TERM (1995-1997)	Freq	TERM	Entropy	TERM	Offset
terrorism	8	forensic science	3.332	avoidance symptoms	7.995
blast overpressure	5	terrorist attack	3.327	early intrusion	7.995
terrorist bombings	5	injuries	3.303	injured survivors	7.995
posttraumatic-stress-disorder	5	warfare	3.286	predictive value	7.995
injuries	4	sarin	3.258	neuropathology	7.909
explosion	4	experience	3.187	warfare	7.860
casualties	4	outbreak	3.142	rat	7.791
sarin	4	casualties	3.100	divorce	7.670
soman	4	identification	3.098	terrorist attack	7.460
organophosphate	4	media	3.093	collective violence	7.388

Figure 4. A network of keywords in the terrorism-research literature (1995–97). High-frequency terms appear in black, whereas outlier terms identified by informational bias appear in dark red.

Figure 4 illustrates how to facilitate a sense-making process with both high- and low-profile patterns embedded in the same visualization. The network in Figure 4 consists of keywords that appeared in 1995, 1996, and 1997, corresponding to the first period of substantial change in terrorism research. High-profile patterns are labeled in black, whereas low-profile patterns are labeled in dark red. High-profile patterns help us understand the most salient terrorism-research topics in this time period. For example, terrorism, posttraumatic-stress disorder, terrorist bombings, and blast overpressure are the most salient. The latter two topics are closely related to the Oklahoma City bombing event, whereas posttraumatic-stress disorder is not directly connected at this level. In contrast, low-profile patterns include avoidance symptoms, early intrusion, and neuropathology. These terms are unique in reference to other keywords. Once analysts identify these patterns, they can investigate even further and make informed decisions. For example, they might examine whether an unexpected topic is appearing for the first time or a new trend is emerging.

Conclusion

Developing methods and principles for representing data quality, reliability, and certainty measures throughout the data transformation and analysis process is a key element on the research agenda for visual analytics.[3] This introduction to an information-theoretic view of visual analytics illustrates its potential contributions to facilitate analytical reasoning and visual analytics. Asking the right question is the key to connecting the dots in visual analytics. Examples in this article show various ways to apply information-theoretic theories, strategies, and techniques to find the dots, make sense of them, and differentiate them at different levels of abstraction—from macroscopic to microscopic. The information-theoretic perspective provides a potentially effective framework to address questions concerning analytical reasoning with uncertainty, synthesizing evidence from multiple sources, and developing a macroscopic understanding of a complex, large-scale, and diverse body of information systematically. It should stimulate further advances in visual analytics and work harmonically with other approaches to facilitate analytical reasoning. ■

Acknowledgments

I am grateful for insightful comments from Alan M. MacEachren and anonymous reviewers on an earlier version of this article. The work is supported by the National Science Foundation under the grant IIS-0612129 and the Northeast Visualization and Analytics Center (NEVAC).

References

1. T. Anderson, D. Schum, and W. Twining, *Analysis of Evidence*, 2nd ed., Cambridge Univ. Press, 2005.
2. M. Gladwell, "Open Secrets: Enron, Intelligence, and the Perils of Too Much Information," *New Yorker,* 8 Jan. 2007, http://newyorker.com/reporting/2007/01/08/070108fa_fact_gladwell.
3. J.J. Thomas and K.A. Cook, *Illuminating the Path: The Research and Development Agenda for Visual Analytics,* National Visualization and Analytics Center, 2005.
4. S. Kullback and R.A. Leibler, "On Information and Sufficiency," *Annals Mathematical Statistics,* vol. 22, no. 1, 1951, pp. 79-86.
5. E.S. Soofi and J.J. Retzer, "Information Indices: Unification and Applications," *J. Econometrics,* vol. 107, no. 1, 2002, pp. 17-40.
6. P. Pirolli, *Information Foraging Theory: Adaptive Interaction with Information,* Oxford Univ. Press, 2007.
7. J. Gill, "An Entropy Measure of Uncertainty in Vote Choice," *Electoral Studies,* vol. 24, no. 3, 2005, pp. 371-392.
8. L. Bartels, "Issue Voting Under Uncertainty: An Empirical Test," *American J. Political Science,* vol. 30, no. 4, 1988, pp. 709-728.
9. L. Itti and P. Baldi, "A Principled Approach to Detecting Surprising Events in Video," *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (CVPR), IEEE CS Press, 2005, pp. 631-637.
10. R.S. Burt, "Structural Holes and Good Ideas," *American J. Sociology,* vol. 100, no. 2, 2004, pp. 349-399.

Contact author Chaomei Chen at chaomei.chen@cis.drexel.edu.

Contact editor Theresa-Marie Rhyne at tmrhyne@ncsu.edu.