

Optimizing the value of computer networks
by
Networking Information
using
Affinity Directories

Thomas L. Bascom
LinkSpace, LLC

Published December 1, 2002 in The Journal of the Knowledge and
Innovation Management Professional Society

Introduction

The World Wide Web and other networks provide access to a vast amount of information. More information than can be viewed, comprehended or absorbed by the individual user. This information glut raises questions regarding the value for users of the collection of information on these networks. This study discusses the value of and motivations for a new technique, “networking information,” to optimize the value of networks. The proposed technique of networking information uses affinity directories that store and distribute the relationships or interconnections among information that is on computer networks and that matches an affinity user’s interest.

Throughout the study, we discuss affinity directories and affinity groups. As used here, an affinity group is a group of people that have in common an interest in a subject. An affinity group may or may not be organized and an individual may belong, explicitly or implicitly, to one or many affinity groups. As used here, affinity directories are classified listings of references to information on computer networks. Affinity directories currently exist. Portals, which are used throughout the Web act as affinity directories and search engines might be considered to be ad-hoc affinity directories. While portals and search engines aid affinity groups in comprehending and accessing available information, this study proposes that there remains opportunity for greatly increasing the value of computer networks by “networking information” on them using a new form of affinity directory. This paper is intended only to introduce the concept. Empirical validation of the proposed improvements in “value” and “satisfaction” gained by networking information is required.

The Value of Networks is understood

Robert Metcalf, the inventor of the Ethernet, proposed “the value of a network can be measured by the square of its number of users. More simply: Connected computers are better. Having the only telephone in the world would be of zero value, but this value increases for each new telephone it can call.”¹ This concept can be represented mathematically as:

$$V_N \propto N^2$$

where V_N is the value of the network and N is the number of people or devices on the network.

This belief that, as a network expands, its value expands exponentially does not take into account the time and comprehension limits of the human user. Leading indexes of the World Wide Web contain over 3 Billion pages. Corporate networks can be poorly organized and indexed making information on them inaccessible.

Dr. David Reed, former vice president and chief scientist of Lotus Development Corporation points out that Metcalf’s law does not capture the power of networked

¹ Bob Metcalf “There ought to be a law”, New York Times 7/15/1996

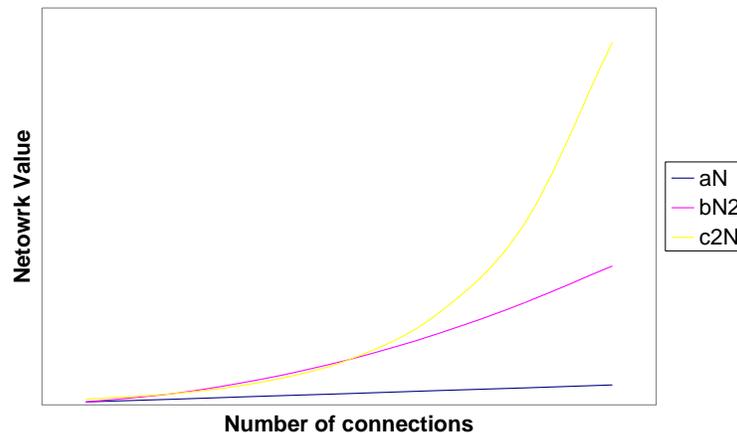
groups. Reed acknowledges that the Metcalf connections are only “potential connections between a pair of customers.”² Reed proposes in his third law that “As networks grow, value shifts: Content (whose value is proportional to size) yields to Transactions (whose value is proportional to the square of size), and eventually Affiliation (whose value is exponential in size)...the "option" to affiliate in groups is also a form of value, and that the set of all subsets of a set has cardinality 2^n , which grows a lot faster than the square [Metcalf] law.”³ Restated by Reed in an interview:

“So any system that lets users create and maintain groups creates a set of group-forming options that increase exponentially with the number of potential members. And as a function, 2^N dominates N^2 - which means that even if each individual group-forming option is worth much less than an individual pair-wise connection, eventually the total set of group-forming options will have far more option value than the pair-wise options.”⁴

Combining the power of Metcalf’s and Reed’s laws then leads to the conclusion that the value of the network is proportional to the number of nodes (connected computing devices) and how they are connected. Reed provides the following value formula and illustration of the contributions as follows⁵:

$$V_N = aN + bN^2 + c2^N$$

Value of the Network



Where a, b and c are constants.

² Dr. David P. Reed “That Sneaky Exponential—Beyond Metcalfe’s Law to the Power of Community Building” a supplement to an article in Context magazine published in Spring 1999, <http://www.reed.com/Papers/GFN/reedslaw.html>.

³ Dr. David P. Reed Reed’s 3rd law: a scaling law for network value, <http://www.reed.com/reeds3rd.htm>.

⁴ Journal of the Hyperlinked Organization interview with Dr. David Reed on 1/19/2001, <http://www.hyperorg.com/backissues/joho-jan19-01.html#reed>.

⁵ Ibid. Reed “That Sneaky Exponential—Beyond Metcalfe’s Law to the Power of Community Building”

The Value of Information On Networks has not been realized

Expanding on Reed's work, Kwak and Fagin, peer to peer industry analysts at Bear Stearns, point out that Reed did not consider possible relationships between [data] objects on the network. Data objects are all of the individual pieces of content on the networks, hereafter referred to as content objects. If the relationships between content objects are considered, the network value equation (without modifying Metcalf's law for the same consideration) becomes:

$$V_N = aN + bN^2 + c2^{NB}$$

where B is the number of objects on the network.⁶ Taking relationships into account, the theoretical value of the network soars as objects on the network are connected. Kwak and Fagin however did not address how these content objects would be connected in groups as is required for Reed's law to apply. Currently content objects are indexed but not bi-directionally connected in the fashion that computing devices are. The ability to realize the value exponent for content objects further requires bi-directional connection as Kwak and Fagin stipulate for Metcalf's law.⁷

Much work is being performed in the area of identifying and using relationships among content objects. These efforts support the value of the networking information.

A well known example is Google™, which works by ranking pages based on their hyperlinks.

“Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weigh more heavily and help to make other pages "important."

Important, high-quality sites receive a higher PageRank....Google combines PageRank with sophisticated text-matching techniques to find pages that are both important and relevant to your search.”⁸

Google™ has generated tremendous value by taking advantage of the existing method of networking of content, simple hyperlinking, to enhance the presentation of search results. This value is demonstrated by their popularity.

Google™ uses a class of product called an “intelligent agent” to interpret this network of hyperlinks. The International Journal on Cooperative Information Systems⁹ describes the following applications for intelligent information agents:

⁶ Chris Kwak, Robert Fagin “Bear Stearns Internet Infrastructure & Services (Introducing Internet 3.0)”, 5/2001 Page 112 Available at <http://wpinter1.bearstearns.com/supplychain/infrastructure.htm>.

⁷ Ibid. Chris Kwak, Robert Fagin pg. 32 Exhibit 12.

⁸ “Our Search: Google Technology”, <http://www.google.com/technology/index.html>.

- *Information acquisition and management*, i.e., retrieve, extract, analyze and filter data
- *Information synthesis and presentation*, i.e., integrate heterogeneous data and provide unified (and multi-dimensional) views on data.
- *Intelligent user assistance*, i.e., provide convenient individual interactive assistance including recommending sources and future work steps, etc.

Any intelligent agent developed for these types of applications certainly might take advantage of the value of networked information demonstrated by Google™.

Currently, however, Google™ relies on the existence of hyperlinks, the direction of the links, and characteristics of the linked content. Clearly, there is opportunity to improve the performance of such intelligent agents by improving the definition of the relationships in the network.

Advancements are being made in the following areas that may help realize the potential of this demonstrated value of networked information:

- Ongoing research and the development of standards, in the fields of taxonomy, ontology, topic maps and epistemology will improve the description of content and perhaps the relationships among content. Work in these areas is focused on the structured assignment (or inference) of meaning for data and for statements made in natural language in order to support indexing, search and re-use, as by intelligent agents.
- The World Wide Web Consortium (W3C) is developing a standard for extended hyperlinking called XLink, which defines styles for hyperlinking, including bi- and omni-directional links. The XLink standard will improve the communication, understanding, and utility of hyperlinks.

An optimized network will contain explicitly defined bi-directional relationships among content objects on that network.

The Value Contradiction of Too Much Information

In January 2000, researchers at NEC Research Institute and Inktomi, Inc. completed a study that estimated that the Web had over 1 billion pages and almost 5 million servers¹⁰. In July 2000, Cyveillance estimated that there were 2.1 billion unique, publicly available pages on the Internet¹¹. Cyveillance states that the Internet grows by 7.3 million pages each day, which is consistent with the two size estimates. These figures represent a vast

9 INTELLIGENT INFORMATION AGENTS: THEORY AND APPLICATIONS International Journal on Cooperative Information Systems, Vol. 10(1&2), March 2001

¹⁰ Inktomi press release, "Web Surpasses One Billion Documents" January 18, 2000.

¹¹ eMarketer, "Size of Net Will Double Within Year" July 11, 2000, http://www.emarketer.com/estats/20000713_size.html?ref=wn.

number of connected computing-devices (N) and content objects (B). The value of access to such vast information resources theoretically becomes infinite. However, it anecdotally creates a condition contrary to the Metcalf and Reed proposition in that, if the potential connections are incomprehensible or un-searchable, the perceived network value diminishes with growth. Even Google™, the most comprehensive web search engine, captures only 42 percent of index-able web pages.¹² In particular, when given a finite time, the perceived value of a search is diminishing as many of the inexhaustible resources are not considered. This relationship between time, volume of data and satisfaction is hereafter described as the “Time to Satisfaction Value” represented as V_{ts} . As the content on the network increases, the Time to Satisfaction Value of the network diminishes. In other words, the user’s satisfaction that they have discovered the content on the network that is relevant, authoritative and complete diminishes as the awareness of alternative resources explodes. The Time to Satisfaction Value may be represented as:

$$V_{ts} \propto 1/(NB)$$

On a single network computing device with one content object, the value is maximized. The user is completely satisfied that they have considered all of the available resources. As the number of objects on the network approaches infinity the time to satisfaction value approaches zero.

Reed proposes that the user may find value in affinity groups on networks that accommodate groups;

“[L]ike the Internet, Group Forming Networks (GFNs) are an important additional kind of network capability. A GFN has functionality that directly enables and supports affiliations (such as interest groups, clubs, meetings, communities) among subsets of its customers. Group tools and technologies (also called community tools) such as user-defined mailing lists, chat rooms, discussion groups, buddy lists, team rooms, trading rooms, user groups, market makers, and auction hosts, all have a common theme—they allow small or large groups of network users to coalesce and to organize their communications around a common interest, issue, or goal”.¹³

Note that these examples connect groups of people (through their devices), not information.

Thus, the Time to Satisfaction Value of a network may be improved for increasing values of N where the user has a priori knowledge that through the group, they are connecting to network elements that contain high affinity to the desired object. A classic example of this benefit is the collaborative filter used by Amazon.com, Inc. to intelligently recommend to visitors items that other visitors who have purchased their current selection have purchased. In other words, the affinity group can increase the value of the network

¹² David Lake. “Engines Idling Roughly,” *The Industry Standard*, February 9, 2001

¹³ *ibid.* Reed “That Sneaky Exponential—Beyond Metcalfe's Law to the Power of Community Building”

to the group's members by accumulating the knowledge or experiences of the group and enabling members to be aware of it. Increasing the possibility that of all the items on the network, the group has considered many more than any individual in the group. For a large group with high affinity, a user might have great confidence that the group has accumulated knowledge of most networks and content valuable to the group. Thus sharing knowledge awareness of the affinity group increases each user's time to satisfaction value for the network as follows:

$$V_{ts} \propto 1 / [(N - N_{A0})(B - B_{A0})]$$

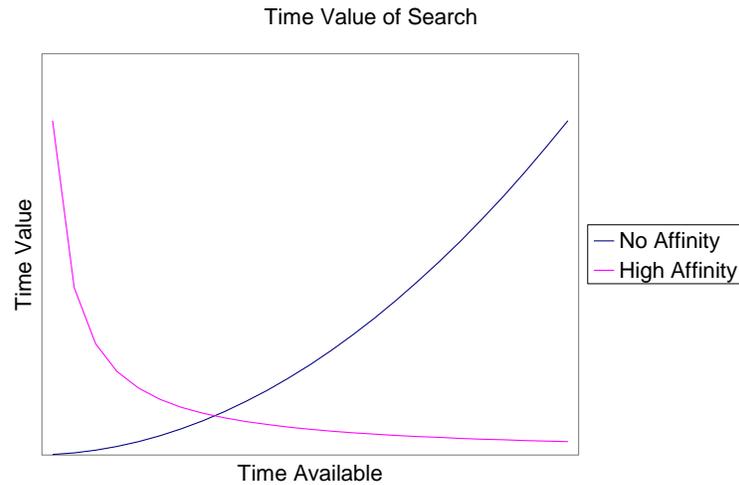
Where N_{A0} represents devices and B_{A0} represent content objects on the network that have no affinity to the group's intent. Alternatively, because there will be degrees of affinity, rather than rejecting connections that have no affinity, the user may seek items with high affinity. Thus, the Time to Satisfaction Value becomes:

$$V_{ts} \propto 1 / (N_A B_A)$$

Where N_A and B_A represents network connections and content objects respectively with a threshold affinity that is satisfactory to the user. A user might select a threshold affinity for their network connections in order to access a finite subset of the network while still satisfying themselves that the available content is important and presented in an ordered format.

Currently, there are affinity directories on existing networks in the form of portals. As mentioned previously, these portals do not network (connect) the information, they only cluster it and present it hierarchically. In addition, the portal market is contracting through consolidation and closures. This creates large, commercially viable, and popular portals, indexes, directories and other community tools containing diverse content with low affinity. Fortunately, with the use of meta-tags, XML and more intelligent search engines, the mechanisms for indexing distributed content help with their affinity searches. However, the persistent user will still have to use search engines to find high affinity portals or affinity content one item at a time and the user will have to review each item for content matching particular search criteria.

There exists a need for the creation and proliferation of economically feasible and accessible affinity directories. These affinity directories can capture awareness of content objects of large groups across the network. These affinity directories can also store attributes of each content object in the directory; attributes that will define their applicability for the affinity directory users allowing the users to achieve fast Time to Satisfaction Value searches by accessing content in the directory that fit the affinity attributes selected by the user. An index with high affinity related to a very esoteric subject will then contain a relatively small number of references to content objects and the Time to Satisfaction Value (V_{ts}) will still be very high. Alternatively, for low affinity directories the V_{ts} will be very low unless there is a great deal of time available. Illustrating V_{ts} versus time available for searches on networks with low and high affinity directories:



The contribution to the network’s value due to the number of affinity directories is characterized as the Affinity Index (A). The affinity index increases with the number of affinity directories. Modifying Reed’s value equation then with addition of the affinity index in consideration of the Value contradiction of too much information and its impact on V_{ts} , the equation becomes:

$$V_N = A (aN + bN^2 + c2^N)$$

The value of a network then improves proportionately to the number of affinity directories or indexes of the content.

Another way to appreciate the value of affinity is from the perspective of the value of a time constrained search. Given a limited time, a user can only consider a limited number of content objects. This relationship between time, and satisfaction is hereafter described as the “Time-Constrained Value” or V_{tc} . The V_{tc} is calculated here as the number of content objects considered (B) times their affinity to the search purpose (a) divided by the time available.

$$V_{tc} \propto B a / t_{avail}$$

The optimized network; networking information in the affinity directory

The intent of the affinity directory is to accumulate the knowledge or experiences of a group and make all members aware of it. An affinity directory structure may also allow the content objects to be bi-directionally connected (networked) themselves. Thus, a network with a combination of affinity directories networking content objects for groups of users approaches an optimized network.

The value of the optimized network with affinity directories that network content objects then becomes:

$$V_N = A (a N + b N^2 + c 2^{NB})$$

This equation again does not consider modifying Metcalf's law in consideration of the connection of content objects on the network.

Economic value of affinity networked (bundled) content objects

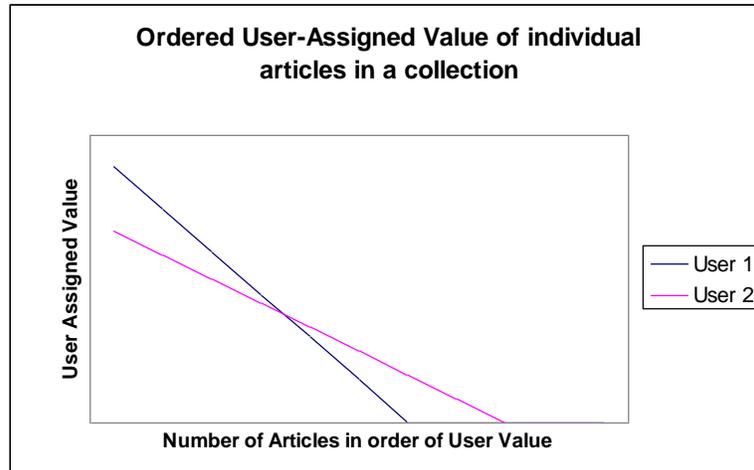
The affinity directory and content object networking value is demonstrated, economically speaking, by their functional similarity to research into the value of publisher's bundling their content as discussed in the work of [Jeffrey K. Mackie-Mason](#)¹⁴, the Arthur W. Burks Professor of Information and Computer Science, and a Professor of Economics and Public Policy at the University of Michigan. Mackie-Mason studied the electronic sale or distribution of content by publishers and the impact of the bundling of content on consumption.

Mackie-Mason proposes that given a collection of publisher's articles, an individual user views one article as the most valued and other articles in the collection as having decreasing value, with some articles having no value at all to the given user. The most valued article, that article's relative value, and the number of articles with value > 0 in a collection varies between users. That is represented mathematically as

$$V_n = V_0 [1 - (n / (KN))]]$$

Where V_n is the user assigned value of article n , V_0 is the user assigned value of the most valuable article in the collection, K is the fraction of articles in the collection with a value greater than zero, and N is the number of articles in the collection that the user has assigned a value greater than zero. Graphically illustrated (for 2 users) this appears as follows:

¹⁴ Jeffrey K. Mackie-Mason "Pricing and Bundling Electronic Access to Information" <http://www-personal.umich.edu/~jmm/presentations/columbia-oct98.ppt>



Assuming the perspective of the Internet as a publisher, and all of the content on the internet as “articles” from that publisher, the same principle applies. For each user, one article is most valuable, others can be ordered with descending value, and many things have no value at all. For each individual, the most valuable, the descending value order, V_0 and K are different. Even for an individual, these values can change over time.

Mackie-Mason also points out that publishers can maximize profits with bundling because information is infinitely flexible in bundling, and because users can be compelled to buy bundles of articles where they wouldn't buy a general subscription or individual article.

Mackie-Mason segregates users as either heterogeneous or homogenous. These categories are analogous to classification of content consumers in this paper as either having low affinity (heterogeneous) or high affinity (homogenous). His work proposes the most profitable bundling choice is based on the cost of the content and the user's homogeneity (affinity). However, a publisher can maximize profits by offering customer selection of bundling options from the following categories:

- General subscription (access to all articles);
- Unbundled (purchase articles one at a time);
- Publisher bundled based on the publishers assessment of the market's interest;
- Bundled based on the users selection of preferences.

His work illustrates the marginal value of subsets of high affinity information to the user. This analogy supports the value of affinity directories serving groups. By bundling access to Internet content in an affinity directory that serves a particular affinity group, that group that is frustrated by broad search engine returns or portal content, and that is not satisfied by content of a few authoritative sources can now be served by the content bundled in an affinity directory.

Networks that include many affinity directories that support networking information fulfill the economic advantage of bundling on the Web by allowing the creation and

selective access of networked content objects on many affinity directories. These directories further allow the user to customize their bundle by opting into directories, and to further select attributes for the content they wish to be offered.

Conclusion:

This study proposes several advantages of optimizing the value of computer networks by networking information on them using affinity directories.

The study proposes that an optimized network will contain explicitly defined bi-directional relationships among content objects on that network.

The study proposes that the optimized network of the future will support affinity directories that support the networking of content objects. The ability to easily create, find, join and contribute to affinity directories and to bi-directionally connect content objects and be made aware of those connections will optimize the network.

The study proposes that the value of affinity directories is demonstrated by their ability to improve a user's:

- Satisfaction with a time constrained search (Time Constrained Value V_{tc}).
- Productivity given affinity ordered subsets of available Content Objects (Time to Satisfaction Values V_{ts}).
- Ability to bundle the most desired content from large collections with attributes selected by the user.

Subsequent research in this topic should include empirical validation of these perceived values