Towards A New Information Theory

Yi-Cheng Zhang*

March 2004

The modern information theory is established by Claude Shannon around the middle of the past century. The key challenges at the time were how to ensure a reliable transmission of signals. A prototype example is enough to illustrate the point. A signal represented by a binary sequence is to be transmitted over a noisy channel, the channel may randomly flip the bit with a given error rate. At the receiver side we'd like to recover the original message correctly, what to do? Since there is no miracle the only way to surmount the difficulty is to send the original bits more than once, in the hope the receiver can figure out the correct original bit. If resource is infinite, the solution is simple: repeat each bit infinite number of times and a simple average on the receiver's side suffices to recover the exact original message. Infinite resources, however, never exist in reality, the work by Shannon and his disciples was to find the minimal necessary number of redundancies in order to recover, or decode the correct signal,if the resources are below this minimal requirement, what best approximation can one obtain. Shannon's Information theory provides a general theoretical framework to construct the most efficient information filter for noisy signals.

With the current rapid advances in information technology, especially with the advent the Internet, there is much more available information for people to be able to reliably select what is relevant and important for them. To cope with such an 'information explosive growth', search engines play a pivotal role. Current the most popular search engine is Google, which demonstrated in a brief time span how much difference a more powerful information filtering mechanism can make. In this essay I'll outline a theoretical framework akin to Shannon information theory, which tackles the basic mechanism

---
*Department of Physics, University of Fribourg, CH-1700, Fribourg, Switzerland. yi-cheng.zhang@unifr.ch

behind current and future search engines. To call attention to scientists, especially physicists, that a host of new challenging models demand in depth study.

Let us first consider an idealized, miniature situation that search engines and other information filters must face. For a given niche a group of experts of varying intrinsic qualities evaluate each other. Denote by $x_i > 0$ the ith expert's quality, i=1,2,...,N. Since we cannot know the God-given qualities directly, we must rely on the mutual evaluations among the experts. We further assume that all experts can make errors in their evaluations, but an expert with a higher quality makes less error than another with a lower quality. We further assume that every expert evaluate all other (N-1) fellow experts, we thus obtain a matrix of $x_{ij}, i \neq j$, i.e. x_{ij} represents jth expert's estimate on ith expert's quality. Since no one knows the God-given qualities a priori, we must rely on the "noisy" signal—the experts' mutual imprecise, often contradictory evaluations represented by the matrix x_{ij}. In other words we must design a way to make sense of the noisy signal x_{ij}, try our best to decode the intrinsic message x_i. We shall first attempt to solve this simplest example, then gradually, we shall relax the artificial conditions and render the model more amendable to real applications.

Denote by x_i' our solution, a good solution will yield $\{x_i'\}$ and $\{x_i\}$ very close to each other or even identical. The parallel with the Shannon's problem is striking: for us $\{x_i\}$ is the original message, we strive to get the best decoded message $\{x_i'\}$. The The simple proposal is to take the arithmetic average, $x_i' = c\sum_{j=1}^{N} x_{ij}$, i.e. it is the average evaluation by all the fellow experts on the ith expert. However, though democratic, this simple solution is not very good: for we do know that some experts' evaluation is more reliable than others. If we give all the evaluations an equal weight, we can never find who are better experts—which contradicts our stated task. Therefore we would like to give more weights to the higher-quality experts, but we cannot know who they are. We propose the following set of implicit equations: $x_i' = \sum_{i=1}^{N} w_j x_{ij}, \ i = 1, ..., N$. The normalized weights w_j is a function of all the variables x_{ij}. We further assume that $w_j = f(x_j')$, this ansatz restricts considerably the space of all the possible functions. $f(x)$ is a single variable function that we must choose to achieve the best solution. The simplest weighted average function is linear: $f(x) = x$. Then we have the weight of jth expert proportional to x_j', i.e. the higher is her received evaluation score, the weightier is her evaluation on her fellow experts. We know none of these x_j' separately, but we can solve the N simultaneous equa-

tions: $x'_i = c \sum_{j=1}^{N} x'_j x_{ij}$, where c is the normalization constant. We now have N equations and as many variables, we can readily find a solution. It's easier to use matrix notations. Denote by $\hat{x}$ the column vector, whose components are $x'_1, x'_2, ..., x'_N$; X the matrix cx_{ij}. The set of equations now stands $\hat{x} = X\hat{x}$. The homogeneous equations do not fix the absolute values of x'_i. The eigenvalues and eigenvectors of X yield both solutions and solution convergence.

The linear weights are considerably easier to handle than nonlinear ones. Moreover, we do not know a priori how non-linear our weights should be. In principle we expect when the experts' qualities vary greatly, we need very non-linear weights to be able to pick the extraordinary contributions from a small group of elites; or when their qualities vary not much the linear function or even a constant weights are more appropriate. Worst of all we don't even know what kind of distribution behind such quality variation. With so much unknowns what can we do? In theoretical model studies we can propose some simple distribution to get experience. For instance we can assume the noisy matrix elements to be $x_{ij} = x_i \exp \epsilon / x_j$, ϵ is randomly drawn from -1 to +1. In real situations we can proceed as follows: start out with a given initial configuration for $\{x'_i\}$. Using the set of the equations $x'_i = \sum_{j=1}^{N} f(x'_j) x_{ij}$ iteratively to find a stable solution. If starting from a set of different initial configurations we always end up in the same solution then we may conclude that there is a finite range of convergence around this solution, likely this is our best solution. Convergence always happens with weak nonlinearity, i.e. $f(x) = x^\alpha$ for α small. Numerically one can start from $\alpha = 0$, gradually increase it. For each value of α we need to check the convergence criterion, until for a high enough value solutions appear to be scattered around with different initial conditions. If we assume the existence of the attractor od attractors, analytical insight can be obtained by studying attractor's stability. Probing with infinitesimal perturbations we find a set linearized equations, the accompanying eigenvalues and eigenvectors yield much information on the nature of these attractors. Though deeper theoretical insight is not yet complete and deserved further study, the above scenario seems to provide a general workable framework.

For small communities the above model can be of use for sorting out ranking problems in small to middle sized communities. For large networks like the world wide web we cannot assume every web site is directed to every other. Also in large communities like all the physicists in the world, each of us knows only a tiny fraction of all others. The above matrix is

likely to be a very sparse one. For sparse matrices in general an information filter works less well, relatively speaking, but there is still sufficiently large number of redundancies to exploit. We can either use the above model, to find best solution iteratively, or by a simpler method, that is inspired by Google's search algorithm. Instead of a community of experts, now we consider the WWW. Each web site designer decide to which his site should point to. Out of the total available web sites it only points to a tiny fraction of them. Pointing to a site represents a measure of approval, though in an very approximate sense. Yet Google is able to capture this input data to render the vast web much easier to navigate around, than its predecessors. Denote by $m_{ij} = 1$, or 0 as the connection matrix. $m_{ij} = 1$ implies that ith site points to jth site, 0 (for the vast majority sites) otherwise. Note that this matrix is not symmetrical as the directed links are in general not reciprocal. A site will be considered important if it is pointed to by sites which themselves are also important. We face a similar problem as the above evaluation problem. Due to the information input is in binary format $0, 1$, we can develop a simpler method to solve it. From the above connection matrix we can construct a metric matrix: for a given pair of sites i and j we can define a directed distance d_{ij}, which is the number of steps along the shortest directed path from j to i. For instance, if the site A points to site B, then B to C, the directed distance from A to C is two, or $d_{CA} = 2$. Therefore out of the matrix $\{m_{ij}\}$ we obtain a new matrix $\{d_{ij}\}$ whose entries are nonzero, positive integers except for the diagonal elements. We can define the importance of ith site as $I_i^\alpha = \sum_j \alpha^{d_{ij}} = \sum_{d=1,2,\ldots} N(d)\alpha^d$, where $N(d)$ is the number of the sites exactly at the directed distance d. This definition values those near sites more than those distant sites. With every site's importance easily tabulated, one can easily find who is relatively more important. The weight parameter $0 < \alpha < 1$ can be fine-tuned to adjust to the task at hand: if one need to value more near neighbors than distant sites α can be made small; if one need to sample relatively large region α should be large. Its value should be empirically decided. The scheme can be easily made to apply to evaluate the impact of scientific papers. A paper refers to a list of other papers, which in turn refer to still many others. Direct reference is more important, whereas indirect references should also be counted, as they determine the importance of the papers which refer to the first paper. Currently we often use Citation Index by ISI as a measure of a scientist's impact. This parameter only counts the direct references at the distance one. This method already is better than the old tradition of counting the total number of papers by a scientist,

4

ignoring completely the reference connections. In a foreseeable future we may see the composite impact factor being used to evaluate research impact.

The above simple examples are aimed to show that an emerging branch of Information Theory presents both new opportunities and challenges for scientists. Physicists are especially equipped to tackle many core issues of the information filtering mechanism, since the tools from the advanced statistical mechanics developed in the last few decades allow us to more effectively deal with large networks. The issues can be typically characterized as follows. Many information sources emit noisy signals, for any single subject, there are multiple opinions and many experts behind them. To best make informational sense of the noisy, redundant information, we can leverage the connectedness of a network, to arrive at a best or most reasonable approximation to the underlying truth. The knowledge, so to speak, resides not in isolated skulls but among the intertwined network, the whole can be larger than the sum of the parts. However, this collectively coded knowledge is not easy to read off, careful, innovative methods must be designed to decode the community's message. The tasks and implications of this new Information Theory cannot be overstated, both for theoretical understanding and applications in our economy and society.

This essay is based on a talk given to the Statistical Physics Conference Yangzhow, China. Much work is carried out at the Fribourg Interdisciplinary Physics Group, in collaboration with Lionel Moret, Paolo Laureti, Hassan Masum, Yi-Kuo Yu, and will be reported in details in forthcoming publications.