Markov Lie Monoid Entropies as Network Metrics

Joseph E. Johnson, PhD
University of South Carolina

A network of N nodes can be exactly described by a matrix of N^2–N non-negative off-diagonal values representing the connection weights among the N nodes. When a network is large and changing every second such as the Internet, then the resulting system has tens of millions of values every second. We have found a method for reducing this vast data into a few (2N and fewer) representative values (network entropy spectral functions, or metrics) in order to track the changing topology for attacks, failures and malicious processes.
Our previous work showed that the general linear group, of transformations that are continuously connected to the identity in n dimensions GL(n,R), can be decomposed into two Lie groups[1]: (1) an n(n-1)-dimensional Markov-type Lie group M(n) that is defined by preserving the sum of the components of a vector, and (2) the n-dimensional Abelian Lie group, A(n), of scaling transformations of the coordinates. With the restriction of the Markov-type Lie algebra parameters to non-negative values, one obtains exactly all Markov transformations in n dimensions that are continuously connected to the identity. More precisely, this system is now a Markov Monoid (MM) as it is a group without an inverse.

In our current work we show that every network, as defined by its connection matrix C_{ij}, is in one to one correspondence to a single

element of the MM Lie algebra of the same dimensionality. It follows that any network matrix, C, is the generator of a continuous Markov transformation that can be interpreted as producing an irreversible flow of a conserved substance among the nodes of the corresponding network. The exponentiation of the MM algebra provides a continuous transformation with rows and columns that constitute normed probability distributions that encapsulate the topology of the network in all orders of expansion. This allows Shannon and generalized (Renyi) entropy functions to be defined on the column and row probability distributions. These (2N) generalized entropies (along with derivatives and functions of these entropies) for these Markov transformations become metrics for the topology of the corresponding network encapsulating all of the network topology in a more hierarchical way. Thus we have tightly connected the fields of Lie groups and algebras, Markov transformations, conserved flows, diffusion transformations, and generalized entropies, on the one hand, to network theory and network topology. We are specifically interested in applying these generalized entropies as metrics for the tracking of network topological changes such as one would expect under attacks and intrusions on internets. We will show our experimental results of monitoring these entropy spectral distributions using two internet tracking applications.

1 Introduction

There is a broad spectrum of mathematical problems that involve the general theory of networks and the associated classification, optimization, and potentially even their dynamical evolution. By a network we mean a set of n nodes (points), some pairs of which are connected with a representative non-negative weight or strength of connection. Such a network can be represented by a connection (or connectivity, or adjancy) matrix C_{ij} whose off-diagonal elements give the non-negative 'strength' of the connection between nodes i and j in the network. Often that 'strength' or 'weight' is as simple as a '1' for a connection and a '0' otherwise. A network can be 'undirected' or 'directed' depending upon whether C_{ij} is symmetric or not thus indicating respectively a symmetric or asymmetrical connection between i and j. There may or may not exist a 'metric distance'

between the nodes or, equivalently, positions for the points in a metric space of some dimensionality, such as airports for airline networks, or substations for power or utility distribution networks. It is well known that the classification of different network topologies cannot be accomplished with just the eigenvalue spectra of the connectivity matrix as there are topologically different networks with as few as five nodes that have the same eigenvalue spectra. One root of the network problem is that although the network is exactly defined by the C matrix, there are n! different C matrices that correspond to the same topology because different C matrices result from different nodal numbering orders. Most network problems become computationally intractable for more than a few hundred nodes. The essential point here is that the n(n-1) off-diagonal non-negative values of C uniquely define a network. The n column values are arbitrary at this point and are undefined.

We are interested in seeking useful metrics (functions of the C matrix) for the description of the topology of large networks such as sub-nets of the internet which might have from a hundred to a million nodes, and thus perhaps a trillion connection matrix values. To be useful, the metrics must be (a) rapidly computable (as compared to eigenvalue computations), (b) intuitively meaningful, (c) should holistically summarize the underlying topology with a few variables, (d) ideally would offer meaningful hierarchical expansions providing increasing levels of topological detail and (e) these metrics should be invariant under the permutation group on node numbering and thus reflect the intrinsic topology. We are specifically interested in the information flows of which originating node sends data to which destination node; and we are not initially interested in the underlying physical connectivity topology itself nor are we interested in the specific path which the information traverses nor associated distance metrics. Internet transmissions are extremely dynamic and thus to achieve some form of continuity, we envision constructing the C matrix using a summation of information transfers, or weights, over some time window $t-\delta/2$ to $t+\delta/2$, surrounding a time t as $C(t, \delta)$ thus representing the time evolution of the connection matrix.

Given the number of connections, this problem resembles the representation of a physical gas in terms of thermodynamical variables

(such as temperature, volume, pressure, heat, and entropy). Generally, in such internet environments there is no meaningful location or position metric that gives insight into the topology and thus distance is not usefully defined. As such pressure and volume, do not have a clear meaning without a distance function. There is no general conserved quantity such as energy, and thus heat and temperature do not offer clear meanings. However, we will suggest below that the concept of entropy can be well defined and that it can be used to summarize the order and disorder in the underlying topological structure.

Initially, how to define entropy on the connection matrix is not clear since both Shannon and Renyi entropies are defined as the log of the sum of the powers of the components of a vector, x_i, representing probabilities: $S = c \log_2 (b(\Sigma x_i^a))$ where $\Sigma x_i = 1$ and where a, b, and c are constants. As such these entropies represent the disorder in the underlying probability distribution. The disorder is a maximum with an even probability distribution and is a minimum when all the probability is in one cell with others having a value of zero. But the connection matrix columns or rows cannot be used as probability distributions since the diagonal of C is totally arbitrary. Even if we make some arbitrary choice of the diagonal values of C and normalize the columns, it is not clear what underlying topological 'disorder' we are measuring. Naturally one can take any set of non-negative numbers and normalize them to unity and compute the entropy of the distribution. But without an underlying mathematical and intuitive foundation for the meaning of this distribution it would follow that the resulting entropy calculation is likewise ambiguous. In this work, we utilize our past work on the decomposition of the general linear group in order to gain insight into how one might define these entropy metrics in useful ways that satisfy the requirements a-e above.

Additionally we will utilize definitions of entropy (or equivalently information as negative entropy). The original argument by Shannon was that if the information of two independent systems is to be additive, and if the information is a function of the probability distribution, and since probabilities of independent systems is multiplicative, then it follows that information (or entropy) must be the log of a power of the probability. More precisely beginning with Shannon one has $I = -\log_2(P)$ so that the probability P of a simple two

state system ('1' or '0') is ½ for each thus giving I=1 bit of information. More generally with the work of Kolmogorov and Renyi' one can consider a probability distribution x_i among n cells (i=1,2, … n) with $\Sigma x_i = 1$ as $I = a \log_2(n \Sigma x_i^b)$. In our work below, one can take any of the generalized Renyi' entropies but we will choose a= 1 and b=2 giving $I = \log_2(n \Sigma x_i^2)$. This can be shown to smoothly generalize the Shannon entropy as a boundary condition for two states. For example when $x_1 = 1$ and $x_0 = 0$ (or conversely) then n=2 and I = 1 for maximum information of one bit thus agreeing with Shannon. Then when there is equal probability and thus no information one has $x_1 = x_0 = ½$ thus I = 0. When a probability distribution is flat, the information function above becomes a minimum but when it peaks then the square of the probability becomes much larger and the information increases as the log of the sum of the squares of the values. In the following we will use information and entropy interchangeably as one is the negative of the other.

2 Background on Markov Lie Groups and Monoids

We had previously shown[1] that the transformations in the general linear group in n dimensions, that are continuously connected to the identity, can be decomposed into two Lie groups: (1) an n(n-1) dimensional 'Markov type' Lie group that is defined by preserving the sum of the components of a vector, and (2) the n dimensional Abelian Lie group, A(n), of scaling transformations of the coordinates. To construct the Markov type Lie group, consider the k,l matrix element of a matrix L^{ij} as a basis for n x n matrices, with off-diagonal elements, as $L^{ij}_{kl} = \delta^i_k \delta^j_l - \delta^j_k \delta^j_l$ with i =/= j. Thus the ij basis matrix has a '1' in position ij with a '-1' in position jj on the diagonal. These n(n-1) matrices form a basis for the Lie algebra of all transformations that preserve the sum of the components of vector. With this particular choice of basis, we then showed that by restricting the parameter space to non-negative real values, $\lambda^{ij} >= 0$, one obtains exactly all Markov transformations in n dimensions that were continuously connected to the identity as $M = \exp(s \lambda^{ij} L^{ij})$ where we summarize over repeated indices and where s is a real parameter separated from λ^{ij} to parameterize the continuous evolution of the transformation. In other words $\lambda^{ij} L^{ij}$ consists of non-negative coefficients in a linear

combination of L^{ij} matrics. This non-negativity restriction on the parameter space removed the group inverses and resulted in a continuous Markov monoid, MM(n), a group without an inverse, in n dimensions. The basis elements for the MM algebra is a complete basis for n x n matrices that are defined by their off-diagonal terms.

The n dimensional Abelian scaling Lie algebra can be defined by $L^{ii}{}_{kl} = \delta^i_k \, \delta^i_l$ thus consisting of a '1' on the i,i diagonal position. When exponentiated, $A(s) = \exp(s \lambda^{ii} L^{ii})$, this simply multiplies that coordinate by e^s giving a scaling transformation. The Lie algebra that results from the sum of the Abelian and Markov Lie generators is sufficient to generate the entire general linear group that is connected to the identity.

3 Connecting Markov Monoids to Network Metrics

We can begin with the simple observation that (1) since the non-negative off diagonal elements of an n x n matrix exactly define a network (via C) and its topology with that node numbering, and (2) since a Markov monoid basis is complete in spanning all off-diagonal n x n matrices, then it follows that such networks are in one to one correspondence with the elements of the Markov monoids. The Lie Markov matrix that results is exactly the C matrix where the diagonal elements are set equal to the negative of the sum of all other elements in that column. Thus each such altered connection matrix is the infinitesimal generator of a continuous Markov transformation and conversely. This observation connects networks and their topology with the Lie groups and algebras and Markov transformations in a unique way. Since the Markov generators must have the diagonal elements set to the negative of the sum of the other elements in that column, this requirement fixes the otherwise arbitrary diagonal of the connection matrix to that value also (sometimes referred to as the Lagrangian)

It now follows that this diagonal setting of C generates a Markov transformation by $M = e^{\lambda C}$. One recalls that the action of a Markov matrix on a vector of probabilities (an n-dimensional set of non-negative real values whose sum is unity), will map that vector again

into such a vector (non-negative values with unit sum). The next observation is that by taking λ as infinitesimal, than one can write $M = I + \lambda C$ by ignoring higher order infinitesimals. Here one sees that the value or weight of the connection matrix between two nodes, gives the M matrix element as the relative infinitesimal transition rate between those two components of the vector. Thus it follows that given a probability distribution x_i distributed over the n nodes of a network, then M gives the Markov transition (flow) rates of each probability from one node to another. Thus it follows that the connection matrix gives the infinitesimal transition rates between nodes with the weight reflecting that exact topology.

Specifically, if the hypothetical initial probability vector is $x_i = (1,0,0,0\ldots0)$ then the vector at a time dt later will be equal to the first column of the M matrix, $M = I + dt\ C$. Thus the first column of M is the probability distribution after an infinitesimal time of that part of the probability that began on node 1 and likewise for all other nodes thus giving a probability interpretation to each of the columns of M as the transfer to that node. Thus each column of M can be treated as a probability distribution associated with the topology connected to that associated node and will support an unambiguous definition of an associated entropy function that reflects the inherent disorder (or order) after a flow dt. Thus the columns of M support a meaningful definition of Renyi entropies which in turn reflect the Markov transformation towards disorder of the topological flow to the node for that column. Thus this Renyi entropy on this column can be said to summarize the disorder of the topology of the connections to that node to that order of the expansion. It follows that the spectra of all nodes reflects in some sense the disorder of the entire network. We recall that the numbering of the nodes is arbitrary and thus we can now renumber the nodes without affecting the underlying topology We thus sort the N values of the nodal entropy in descending order which gives a spectral curve independent of nodal ordering and thus independent of the permutations on nodal numbering (except possibly for some degeneracy which we address below). That spectral curve can be summarized by the total value for the entropy of all columns (since entropy is additive and the column values are totally independent

If the connection matrix is symmetric then the graph (network) is said to be undirected, but if there is some asymmetry, then the graph is at least partially directed where the flow from i to j is less or greater than the converse flow. If the connection matrix is not symmetrized then one can capture this asymmetry by resetting the diagonal values of C to be equal to the negative of all other row values in that row. Then upon expansion of $M = I + \lambda C$, the rows are automatically normalized probabilities that in turn support entropy functions for each row. These row entropy values form a spectrum which could be sorted by the same nodal values (in order) that is used to order the column values. This will result in a different spectral curve that is not necessarily in non-decreasing order for the row entropies. One also can compute the total row entropy as we have done for columns. If two columns have the same entropy then one can remove some of the numbering degeneracy by using the values of the associated row entropies by using a rank ordering as we did with column values.

4 Practical and Computational Considerations

The work here has both purely mathematical and practical aspects pertaining to applications to real networks. If one only has a single C matrix and time is not involved then the following discussion on time windows does not apply. It will then be assumed that one has a data flow with records with the fields: (a) network type, (b) time, (c) node i, (d) node j, (e) weight. These might be SNORT captures of internet traffic between IP addresses, financial transitions between bank accounts, power transfers among electrical grid substations, or passengers flown between two airports. The $C(t,\delta)$ matrix is constructed by summing the weights into the appropriate cells (renumbered with integers as i, j = 1, 2, …N) during a time period δ centered about time t. It is obvious that one must have a period δ which allows a 'representative' accumulation of values for the disagrigation size N. If C is too sparse, then one must choose longer time windows or one must collapse the matrix nodes by some natural methodology such as IP sectors, or flights between states and not airports. In some cases one may wish to combine several network types using a linear combination of the contributions determined by the first parameter. In some considerations, one might wish to modify the

weight of the contribution such as using the log of the financial transfer. The software we have built contains loaders with such adjustable parameters. The result of this process is a C(t) with no diagonal terms. We then put this in the form of a Lie Monoid generator by setting the diagonal terms equal to the negative of the other terms in that column (and later row). We then find it useful to normalize the entire matrix to have a fixed trace of -1 or –N as this allows better control over the subsequent expansion into the Markov matrix.

The expansion $M(t) = e^{\lambda C(t)}$ although mathematically guaranteed to converge, have non-negative terms and generally be Markovian, must be executed with a small number of terms if C is large. The parameter λ gives a weighting of the higher terms in the expansion where one might choose to sum up through 'k' terms. The number of such terms is the extent to which M 'feels out' the connections to the connections etc. as weighted by the parameter λ. These two must work hand in hand since it is meaningless to have a very large λ while only expanding to the first order in C. Conversely, it is meaningless to expand to many powers, k, of C while using a nearly infinitesimal value of λ since higher orders of λ will make such higher powers of C vanish. The next consideration is that although the M matrix has only positive terms when the full expansion is executed, in practice one can choose k and λ which, due to the negative diagonals of C, can give negative terms for truncated expansions. Thus the software must have error checks to make the appropriate corrections in the expansion.

Now having the M(t) matrix for that instant, one computes the $E_j^c = \log_2(N(\Sigma_i M_{ij}^2))$ ie the log of the sums of squares of each column to get the entropy (information) for that column representing the transfers into that node by the Markov matrix. The spectra is computed by sorting these by value while keeping a lookup table for which node goes to which original position. A similar computation is done to compute the entropies of the rows E_j^r where the same sort order is used except for removing potential degeneracies (where the column values are the same and thus not distinguished by order). These two spectral curves, or histograms, are computed for each successive time window and overlaid graphically to compare the row and column entropy

profiles over time. A critical point is to realize that it does not matter that the nodes are renumbered with each window, but rather we are interested in whether the profile of order and disorder of the underlying topology is 'about the same'. Naturally some profiles for networks change from late Sunday night to rush hours at 9AM Monday. Likewise, power grids depend upon the temperature as well as the time of day. Thus for a given time of day, day of week, and if necessary for that network, weather pattern in temperature, one must learn the profile of what is normal (i.e. profile one standard deviation) for the network under consideration and then to overlay the instantaneous network spectra on this and graphically display it. One can sum all of the row entropies into a single value Er(t) and likewise for the columns. Then one might sum the squares of deviations from normal to obtain a single value representing the total deviation of column entropies from normal (and likewise for the rows). Our software performs these computations and displays along with the overall network 'amplitude' which is the trace of the original C matrix. This gives us three curves that we can monitor over time as well as watching the current row and column entropy spectra displayed overlaid upon the normal distribution for those circumstances. One must then be able to identify where anomalies are occurring in the network for example by clicking on the associated spectral curve anomaly area. The system then finds the node identification in the lookup table thus identifying the anomalous nodes and subnets. We will present results of our monitoring internet networks at two universities.

5 Interpretation and Discussion

We emphasize again that the flows that are modeled by $M(t) = e^{\lambda C}$ have nothing at all to do with the dynamical evolution of the network. These metrics are used to monitor the network state and dynamical behavior but not to predict it. Rather the evolution generated by $M(\lambda)$ is an imaginary dynamical flow that would occur if a conserved fluid (probability, money, population …) were to move among the nodes at the rates indicated by the C matrix of connected weights. Thus the value of $M(\lambda)$ is that the associated entropies can be used to summarize the order or disorder of the incoming or outgoing topological connectivity of the (static) network at one given instant of time. The

philosophy here is that the entropy values will capture the most essential aspects of the structure of the column and row probability distributions, and thus the topology, to that level of expansion of the parameter λ. By expanding to higher powers of C, with larger values of λ, the entropy metrics capture increasing levels of the connections to the connections etc. Also by utilizing other Renyi' entropies, one obtains other spectra and values that measure other 'moments' of the probability distributions.

One can also consider alternative diagonal values of the C matrix by adding the Abelian scaling group transformation generators to the diagonal values of C. These transformations destroy the conservation of the modeled flow (such as probability) and thus the resulting transformation is no long Markovian. These altered diagonal transformations are equivalent to adding sources and sinks of the modeled fluid at the associated nodes. It is straight forward to prove that the entropy value $E(t) = \log_2(N\langle x(t)|x(t)\rangle)$ when taken to only the third level of expansion, can, with its partial derivatives with respect to such sources and sinks at the node 'j', for different initial conditions for the flow $|x(0)\rangle$ at node 'i', formally obtain the entire C matrix thus showing that the entire topology of the network is contained in the entropy functions and its derivatives.

When C is diagonalized, with the values leading to the Markov transformations, or to the more general values of the diagonals of the last paragraph, one automatically gets a diagonalization of the M matrix. The interpretation of the eigenvectors is now totally obvious as those linear combinations of nodal flows that give a single eigenvalue (decrease when the transformation is Markov) of the associated probability, for that eigenvector. This follows from the fact that all Markov eigenvalues are negative except the one value for equilibrium which has eigenvalue unity for equilibrium. That means that each of these negative eigenvalues of C reflect the decreasing exponential rates of decrease of the associated eigenvector as the system approaches equilibrium as λ approaches infinity in $M= e^{\lambda C}$. This insight allows us to see that all of the Renyi entropy values are increasing as the system approaches equilibrium, which is normally the state of all nodes having the same value of this hypothetical probability. The use here of this

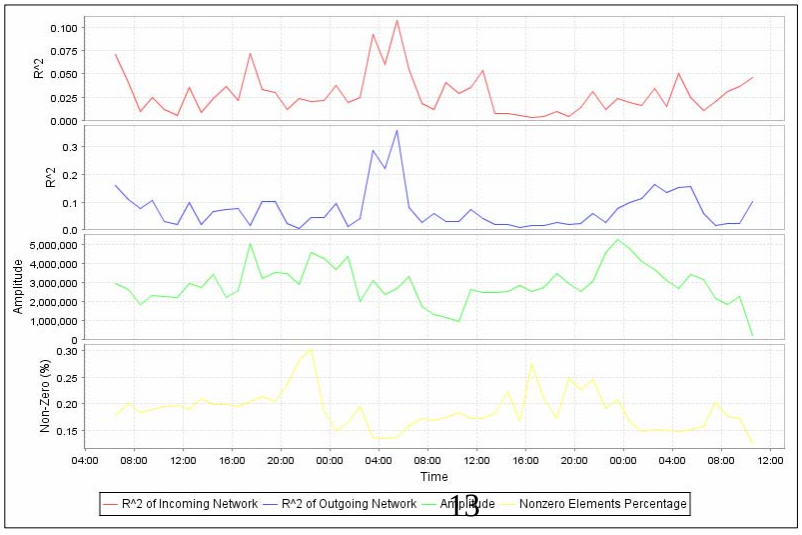
'artificial flow of probability under M' provides us with more than just a method of encapsulating the topology with generalized entropy values, it also gives an intuitive model for the eigenvectors and eigenvalues for C and sheds light on the graph isomerism problem (different topologies having the same eigenvalue spectra). Of course it does not resolve any graph isomerism issue associated with degeneracy of multiple topologies for a single eigenvalue spectra without altering the C matrix by the Abelian transformations.

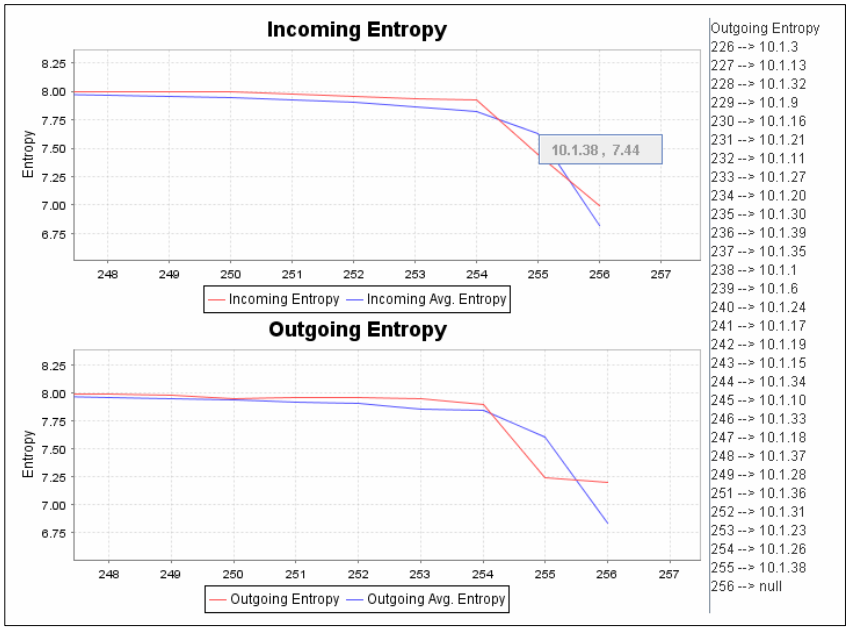
Based upon the arguments above, we suggest that for real networks such as the internet, that the appropriate connection matrix be formed, from source and destination information transfers, where both asymmetry and levels of connection are to be maintained in the C(t) matrix values during that window of time about that time instant. Specifically, this means that if a connection is made multiple times in that time interval, then that C element should reflect the appropriate weight of connectivity as this adds substantial value to the entropy functions. We then suggest that at each instant, the column and row entropy spectra be computed along with the total row and column entropy and that this be done for lower order Renyi entropies as well as lower order values in the expansion of the Markov parameter λ that includes higher order connectivity of the topology. We are currently performing tests to see how effective these entropy metrics are in detecting abnormal changes in topologies that could be associated with attacks, intrusions, malicious processes, and system failures. The patterns (from our simulations) of specific topologies such as rings, trees, clusters, and other structures have interesting entropy spectra. We are performing these experiments on both mathematical simulations of networks with changing topologies in known ways, and also on real network data both in raw forms and in forms simulated from raw data. The objective is to see if these metrics can be useful in the practical sense of monitoring sections of the internet and other computer networks. It is important to note that one can obtain these same metrics for subnetworks of the original network. The subnetwork would be chosen as that portion of the topology that has incoming or outgoing entropy changes that are anomalous. Thus this technique allows an automated reduction or hierarchical expansion methodology to drill into the network to monitor those subnets that are most dynamically aberrant.

6 Results of Monitoring Internet Traffic

The mathematical and computational techniques defined above along with the associated Markov entropy network metrics can be used to analyze the static and track the dynamic behavior of any type of network structure. This includes electrical grids, natural gas pipelines, communications networks, financial transactions, disease networks and social networks. But the network tracking that we have performed to date concentrated totally on internet traffic as defined by Snort data capture at servers of information that is sent from one IP address to another IP address. Our objective is to identify anomalies, and abnormal behavior relative to normal traffic patterns by monitoring the total column (incoming traffic) and row (outgoing traffic) second order Renyi' entropy along with the traffic volume which is independent of the traffic topology. This is similar to separating the buying pattern of financial investments from the volume of transactions on the market as two separate indicators.

The associated graph shows the total incoming and outgoing entropy as a function of time for a server at a university of 30,000 students and faculty. The major anomalies were identified at certain times and these were expanded to see the full entropy spectra at those times over the network thus identifying the specific nodes that had aberrant behavior. It was determined that these particular anomalies in entropy occurred for nodes that at certain times were used to upload and download large volumes of audio and video files.





7 Conclusions

We have proposed network metrics which are generalized entropy functions of the Markov monoid matrix M generated by an altered connection matrix C. When sorted, the associated entropy spectra for the columns and rows of C monitor the state and time evolution of the incoming and outgoing entropy at network nodes. These well defined functions satisfy our original criteria of being fast to compute (compared to eigenvalues), intuitive in interpretation, and hierarchical in revealing sequentially detained network information. They can be used to dynamically monitor networks relative to such normal metrical values thus identifying when the network statistically alters its intrinsic patterns of connectivity.

1. J. Math. Phys. 26 (2) 1985, “Markov-type Lie groups in GL(n,R)”, Joseph E. Johnson

Acknowledgements

Extensive calculations and creation of the associated computer code, Internet data capture, and extensive technical support was performed by Ms. Nayeong Jeong and Mr. John William Campbell. Without their help, the experimental portions of this work could not have been accomplished.