A Network Science Approach to Entropy and Training

Ira S. Moskowitz[1], **William Lawless**[2], **Paul Hyden**[1], **& Ranjeev Mittu**[1]
[1]Information Management and Decision Architectures Branch, Code 5580
Naval Research Laboratory, Washington, DC 20375
&
[2]Department of Mathematics, Sciences and Technology / Department of Social Sciences
Paine College, Augusta, GA 30901

Abstract

In this paper we introduce a new measure of team behavior called *Efficiency* based upon work of Conant on team interdependence. We use Shannon's information as a basis our measure. We then introduce a 2nd law of thermodynamics approach to team efficiency. We conclude the paper by relating our definition of team efficiency to the spectral analysis of the adjacency matrix from the areas of epidemiology and network science.

Introduction

Research within individual and team performance has been an area of significant study for more than ten years. Teamwork is critical within so many domains, for example, within submarine combat information centers, medical, dismounted infantry, virtual and distributed teams supervising multiple autonomous vehicles, etc. Therefore, understanding the processes by which individuals and teams build and share knowledge with other members, techniques to quantify the quality of the knowledge products generated through written artifacts or through discourse, and how these products and processes impact the domain specific mission outcome measures has drawn considerable interest from both academicians and practitioners. This field has received great attention from cognitive scientists and psychologists. For instance, macrocognition (Salas, Fiore, and Letsky 2012, e.g. Ch. 2), which is defined as the internalized and externalized high-level mental processes employed by teams to create new knowledge during complex, one-of-a-kind, collaborative problem solving. High-level processes combine visualizing and aggregating information to resolve ambiguity in support of the discovery of new knowledge and relationships. The field of macrocognition has spawned research that has focused on measuring the convergence of team mental models, metrics that measure the emergence of knowledge and the effects of sharing that knowledge on team outcomes, identification of leaders within a team based on discourse analysis, and the TARGETs (Targeting Acceptable Responses to Generated Events and Tasks) (Stanton et al. 2004, Ch. 53) methodology. This latter approach recognizes that certain team behaviors should occur at certain key events.

While macrocognition and the related fields of study have focused primarily on the cognitive aspects of individual and team performance, very little has been devoted to mathematically modeling team efficiency from the perspective of interdependence, as well as identification of optimal team structures that permit knowledge to be spread effectively. By this, we mean how is team interdependence measured (for those activities that are dependent) and how does team efficiency as measured by the degree of interdependence relate to the spreading of knowledge across the team and eventually what is the impact to improving mission outcome measures. Our work extends the cognitive science based approaches for understanding team performance by applying the notion of joint entropy to understand team efficiencies under conditions of interdependence, and we hypothesize that our technique would enable approaches such as the TARGETs methodology easier to quantify.

In (Balch 1997), (Balch 2000), (Kenny, Kashy, and Bolger 1998), (Lawless et al. 2013), various mathematical approaches are taken to model social interactions. What we find particularly of interest is how a group of individuals is connected into a team in a network of influence. We will model how the team learns its job(s) by using information theoretic measurements such as entropy (Shannon 1948).

This paper does not pretend to be *the* modeling tool for how a team should interact. Rather, this paper, and the ensuing discussion of team efficiency, and the interacting team-graph topology must be taken in the light of desired team interdependencies. In this matter, the paper relies heavily upon the earlier work of Conant (Conant 1974; 1976; Conant and Ashby 1970).

Shannon's seminal work on information theory (Shannon 1948) influenced both Conant and us. We caution the reader, just as Shannon cautioned researchers with information theory in (Shannon 1956), we too must be careful with what we are writing and how it is interpreted. First, we are assuming away the issue of desired team independence. We are only looking at how team players are dependent upon each other. Secondly, when we discuss a closed team, we mean a team without any team degrading interruptions. Thirdly, very similarly to how information theory models how channel capacity lessens following a cascade of channels (Cover and Thomas 2006), we discuss how a team with a high spectral radius (that is much link interconnectivity) makes one team more efficient that a less connected team. With this caveats in mind we continue our paper. We view this paper as a preliminary approach to improving team performance using entropic measures.

A *network* is an undirected graph consisting of nodes (vertices) and links (edges). A *path* between two nodes is a set of links that connect the two nodes. Of course, a link is a path of length one. If we have a node n_A we call all the nodes that have a link to n_B the *neighbors* of n_A, we may also say that they are one hop from n_A. The minimal number of links (minimal path length) between two nodes is called the *distance* between those nodes. For simplicity we assume that the networks in this paper are topologically connected, and that all links are bidirectional and of weight one.

A *team* may be a network in its own right, or a connected subnetwork within a larger network. A team has one or more nodes. We view the nodes as *players*. In an efficient (good) team the players interact well with each other and their actions are highly coordinated with the actions of their teammate players. A team may influence another team by spreading its knowledge across the network in a virus like manner. We desire that a team performs a specific task (it could be reviewing papers for a conference, or making sandwiches at a deli, etc.). The definition of a task is particularly important. A high degree of team player interdependence is desirable (mapping to good performance). However this is not always the case, as certain tasks may require players to be independent. The tasks that involve player independence are outside the model of this paper. This is not a deficit in our model, rather we make the assumption that knowledge shared across a team only helps team interdependency. Tasks that are independent are not (negatively) affected by knowledge sent across the team network. We realize this is a contentious point and will address it further in future work. We refer the interested reader to (Surowiecki 2005, 2.II) for a contrary point of view which is outside of our modeling universe. We desire a mathematical measurement to determine if the team is performing efficiently (well), or if it is performing poorly, based upon *increasing player interdependency*.

Consider the player nodes in a team they are $\mathfrak{X}_1, ..., \mathfrak{X}_m$. At this point, we do not discuss the graph topology for the team, it suffices that the team forms a connected set as a graph. The decision action of each node is given by a random variable $X_k, k = 1, ..., m$ and with probability distribution $P(X_k = x_{k_i}) := p(x_{k_i})$. Each random variable has its associated entropy $H(X_k)$, where[1] as usual

$$H(X_k) := -\sum_{k_i} p(x_{k_i}) \log p(x_{k_i}) \ . \tag{1}$$

Note that for now we are assuming that our distributions are stationary, that is they do not change over time. We may consider the ensemble of team random variables and form the joint entropy (Feinstein 1958, p. 13) and conditional entropy

$$H_{\text{joint}} := H(X_1, ..., X_m) := \\ -\sum_{1_i, 2_i, ..., m_i} p(x_{1_i}, x_{2_i}, ..., x_{m_i}) \log p(x_{1_i}, x_{2_i}, ..., x_{m_i}), \tag{2}$$

$$\text{and } H(X|Y) := H(X, Y) - H(Y) \ . \tag{3}$$

It can easily be shown (Feinstein 1958, p. 16), (Cover and Thomas 2006) that

$$0 \leq H(X_1, ..., X_m) \leq \sum_k H(X_k) \ , \tag{4}$$

with equality iff the X_k are independent.

So, the above tells us that $H(X_1, ..., X_m)$ is maximized when the X_k are independent. Let us look at the opposite extreme, that is $X = X_1 = X_2 = \cdots = X_m$. To do this we use (Cover and Thomas 2006, Thm. 2.5.1) the chain rule for entropy

$$H(X_1, ..., X_m) = H(X_1) + H(X_2|X_1) + H(X_3|X_2, X_1) \\ + \cdots + H(X_n|X_{n-1}, ..., X_1) \tag{5}$$

Since entropy and conditional entropy is never negative we have that

$$H(X_1) \leq H(X_1, ..., X_m)$$

but since we can reorder the X_k without changing the value of the joint entropy (since the commas represent intersection) we have

$$\max_k H(X_k) \leq H(X_1, ..., X_m) \tag{6}$$

Combining the above we have

$$H_{\text{max}} := \max_k H(X_k) \leq H(X_1, ..., X_m) \leq \sum_k H(X_k) \tag{7}$$

In the "most independent situation" we have

$$H(X_1, ..., X_m) = \sum_k H(X_k)$$

and in the most "dependent situation", that is where $\forall k, X_k = X$ we have

$$H(X) = H(X_1, ..., X_m)$$

Thus, the joint entropy give us a metric for how dependent/independent the random variables are, and for our interests for how well the team T is interacting. For a well-functioning team we *want dependence* between the players. So, for our thinking the best way for a team to perform is to be as interdependent as possible, so the closer that joint entropy is to $\max_k H(X_k)$ from above the better, and the closer, from below that it is to $\sum_k H(X_k)$, the worse it is performing. Furthermore, we would like to have this metric of team efficiency normalized so that we can compare different teams.

We propose the following as our definition of *team efficiency* for team T with players modeled by the random variables $X_1, ..., X_m$

$$\mathcal{E}(T) = \frac{\Big(H(X_1, ..., X_m)\Big)^{-1} - \Big(\sum_k H(X_k)\Big)^{-1}}{\Big(\max_k H(X_k)\Big)^{-1} - \Big(\sum_k H(X_k)\Big)^{-1}} \\ = \frac{\frac{1}{H_{\text{joint}}} - \frac{1}{\sum H}}{\frac{1}{H_{\text{max}}} - \frac{1}{\sum H}} \tag{8}$$

[1] Unless noted otherwise all logarithms in this paper are base 2.

The above is well-defined as long as two or more of the random variables are non-deterministic. We see that

$$0 \leq \mathcal{E}(T) \leq 1$$

and when the random variables are independent $\mathcal{E}(T) = 0$, and as the dependence grows $\mathcal{E}(T) \rightarrow 1$, from the left. Keep in mind that 1 is the upper limit.

Examples

Example 1: Say that our team T consists of two players $\mathfrak{X}_1, \mathfrak{X}_2$. Player $\mathfrak{X}_1$ either toasts hamburger rolls or hotdog buns, the behavior of $\mathfrak{X}_1$ is given by the random variable $X_1 = 0$ (hamburger roll), or $X_1 = 1$ (hotdog bun). Player $\mathfrak{X}_2$ either cooks hamburgers or hotdogs. Its behavior is given similarly by the random variable X_2. We want there to be dependencies between the two random variables. In fact, in an ideal world we would have $X_1 = X_2$, that is hamburgers and their rolls and hotdogs and their buns go hand in hand with respect to the production there of. Under that assumption $\mathcal{E}(T) = 1$. If the two players are independent of each other then we have $\mathcal{E}(T) = 0$. Therefore, the better team players we have the closer $\mathcal{E}(T)$ gets to one. (Note: going back to desired player independence, we could have the situation of the cooker of the hamburger and hotdogs needs to add more charcoal to the grill, this action should be independent of what the roll/bun toaster is doing and is not included in our modeling!)

Example 2: Let us consider identical random variables X and Y which describe a fair coin flip. X takes on the values $x_i, i = 1, 2$ and $P(x_i) = 1/2$. Similarly for Y, $P(y_j) = 1/2, j = 1, 2$.

If X and Y are independent (which would make them i.i.d. in this example) we have $p(x_i, y_j) = 1/4$, and $H(X,Y) = 2$. If $X = Y$, which is the most dependent they can be then $p(x_1, y_1) = 1/2 = p(x_2, y_2)$, so $H(X,Y) = 1$. Consider the following matrix which describes the joint distribution

$$J = \begin{pmatrix} p(x_1, y_1) & p(x_1, y_2) \\ p(x_2, y_1) & p(x_2, y_2) \end{pmatrix} = \begin{pmatrix} \frac{1}{4} + \gamma & \frac{1}{4} - \gamma \\ \frac{1}{4} - \gamma & \frac{1}{4} + \gamma \end{pmatrix} .$$

Note that by using J to model the joint distribution the marginals are constant at $p(x_1) = p(x_2) = \frac{1}{2} = p(y_1) = p(y_2)$. and hence $H(X) = 1 = H(Y)$. The joint entropy is

$$H_\gamma(X,Y) = -2\left(\frac{1}{4} + \gamma\right)\log\left(\frac{1}{4} + \gamma\right) - 2\left(\frac{1}{4} - \gamma\right)\log\left(\frac{1}{4} - \gamma\right) .$$

In the above γ is a metric for the dependence between the identical random variables X and Y. When $\gamma = 0$ the distributions are the most independent and $H_0(X,Y) = 2$, when $\gamma = \frac{1}{4}$ they are the most dependent and $H_{.25}(X,Y) = 1$. We see that the efficiency is

$$\mathcal{E}_\gamma(T) = \frac{\frac{1}{H_\gamma(X,Y)} - \frac{1}{2}}{1 - \frac{1}{2}} = \frac{2}{H_\gamma(X,Y)} - 1,$$

which is plotted in Figure 1.

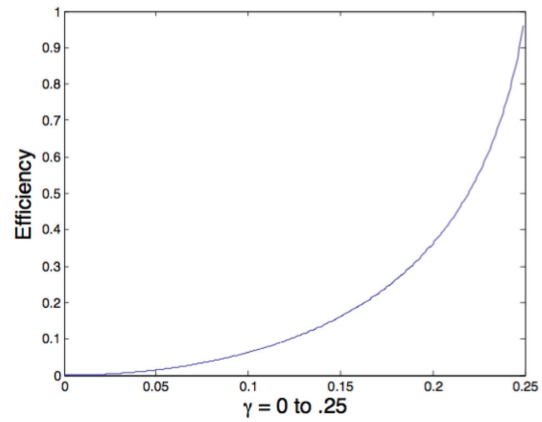

Figure 1: Efficiency

Comments on Shannon and Conant

When we look at Shannon's (Shannon 1948) work on information theory we see that his interest, and a good deal of the information theory community's is on the *mutual information* $I(X,Y)$ between discrete two random variables X and Y, where

$$I(X,Y) = H(X) + H(Y) - H(X,Y) .$$

From our above discussions we see that X and Y are independent if and only if $I(X,Y) = 0$. Mutual information is a measure of how much information inherent in the input X a channel passes through to the output Y. In Shannon information theory the conditional random variable $Y|X = x_i$ is fixed, but we vary the distribution of X to find the one that maximized $I(X,Y)$ and gives us C, the channel capacity.

Our paradigm is opposite to this. We "vary" the conditional values, and do not try to maximize the mutual information, rather we try to maximize our normalized metric efficiency. We note that Conant (Conant 1976) looked at the difference

$$\sum_k H(X_k) - H(X_1, ..., X_m) ,$$

which he coined the *transmission*, this is actually the best we can do for a generalization of the mutual information. We thankfully acknowledge that Conant was the first on the path that we are on now, but we believe we have gone further.

Nota Bene

We make the assumption that we must carefully model *efficient* team work, not the exact opposite. This is due to the phenomena in Shannon's information theory that it is not the actual symbol values that are important but how we can manipulate and send information with those symbols. For example, consider a binary symmetric channel with error probability p. The capacity of this channel is $1 - h(p)$, where $h(p) = -p \cdot \log(p) - (1-p) \cdot \log(1-p)$. Note that $h(p) = h(1-p)$, and that the channel capacity is the same for both p and $1-p$. This is in accord with Shannon's thinking. If the symbols are flipped we get the same information. To make this even simpler if we send "0" and get "0", send "1" and get "1", information theoretically that is the same as sending a "0" and getting "1", and sending a "1" and getting "0".

Correlation makes the distinction between positively correlated and negatively correlated. Anything that is solely entropy based, such as our definition of Efficiency, does not make this distinction. Therefore, one must make the proper modeling assumptions when it comes to choosing the random variable. In the above Ex. 1.1 if instead if we reserve the position of hotdog buns and hamburger rolls, and strive for maximum interdependence we would be doing our selves a disservice. That is, unless we could convince people that hamburgers are meant to be eaten on hotdog buns, and visa versa. Now, this discussion begs the question that the random variable has not changed.

Therefore, we need some concept of an "arrow" that points in the correct direction for our team efficiency if we are in a situation of no initial knowledge. By that we mean, if we start off with the team performing in a manner that we want, increasing the efficiency metric will increase positive team performance. However, if the team starts off in the wrong direction, such as the reversal of buns and rolls from above, the efficiency must first decrease to zero, and then increase in the opposite direction. Therefore, efficiency alone must be tempered with the notion of an arrow or direction as discussed above in the situation where we start off with no idea of our situation. We observe that Conant and Ashby (Conant and Ashby 1970) touched upon this topic, in a slightly different manner, in their analysis of Sommerhoff's work on regulation.

We note that Kenny et al. (Kenny, Kashy, and Bolger 1998) when they model both team independence and dependence rely on correlation. For the remainder of the paper we make the assumption that we start off in a good situation with respect to team performance, and only wish to improve it.

The 2nd Law

In a team we conjecture that $\mathcal{E}$ changes over time. As time progresses the players influence the other players as the team functions more efficiently. Mathematically, this means that the dependency between the random variables X_k increases, which in turn means that the conditional terms in the chain rule for entropy given in Eq. (5) do not decrease over time, they should hopefully increase, but a team should never get worse unless there are outside influences upon it, such as a new player, or something happens to a player.

Therefore, we have that our mathematical terms are temporal, and should be expressed as $X_k(t), H(X_k(t)), ...$, and especially

$$\mathcal{E}(T)_t \ .$$

We may suppress the T part of the notation to make things simpler and, when it is assumed there is a team T involved, express the efficiency of T as a function of t as

$$\mathcal{E}_t \ .$$

Let us return to our previous discussions. For a team where the players learn from each other even though for small time values $\mathcal{E}_\epsilon$ may be close to zero, over time $\mathcal{E}_{\approx\infty}$ should increase and approach one. Now, this approach need not be monotonic, and there may be attracting values along the way. So in fact, a phase space analysis may need to be performed.

Note, a team need not be a closed system, that is players may enter or leave the team, and players' behavior may be influenced by outside forces. With this in mind, we enter the realm of thermodynamics.

Let us consider the Second Law of Thermodynamics (Giancoli 1980 1991, Ch. 15.6). We will not go into the similarities and differences between Shannon entropy $H(X)$ and thermodynamic entropy S. That analysis is best done via statistical mechanics. We have modeled team efficiency in such a manner that it mimics the behavior of thermodynamic entropy. When we are dealing with a closed (isolated) physical system, the change in entropy is given by the ratio of the heat $Q \geq 0$ added to the system at a constant temperature T.

We paraphrase the Second Law of Thermodynamics as

$$\Delta S = \frac{Q}{\mathsf{T}} \geq 0 \ .$$

Thus, for a closed physical system entropy does not decrease over time, in fact for a non-idealized system it always increases.

Now, let us return to social interaction, teams, and team players. In a closed team, when there are not external interruptions, players should learn from each other and their efficiency (if modeled properly!) should not decrease as time increases; hopefully, it should increase. We now state what we call the second law of team dynamics. Note, there is no first law, we are trying to stay consistent with the physics nomenclature.

The Second Law of Team Dynamics:
For a closed team, efficiency does not decrease over time. We take some mathematical liberties writing this as

$$\frac{d\mathcal{E}_t}{dt} \geq 0 \ . \tag{9}$$

Now, what if the team is not a closed system? If new players are introduced, or if players substitute for other players, the efficiency may go down. But, since our system is no longer closed we do not violate the second law.

We say that a team is *non-functional* if $\forall t, \frac{d\mathcal{E}_t}{dt} = 0$. Unless otherwise noted we assume that our teams are striving to improve their performance, never degrade it. Note, this does not mean that $\frac{d\mathcal{E}_t}{dt}$ is never zero, it means that the efficiency is not constant throughout time.

Improving Efficiency

We wish to discuss how the **underlying graph topology of a team can influence its efficiency**. The team is now modeled as a discrete stochastic process. If two players are only one hop away in the graph structure (a direct link between the two players), the random variables can influence each other in the next time step. If they are say, two nodes away, this influence may not occur until the second time increment, etc. Thus team knowledge and dependence flows discretely in time.

As noted, one of our assumptions is that T can be viewed as a connected graph with the players being the nodes and links between the nodes, or that the graph structure can model communication from one team to another. Players can directly influence another player if there is a link between them; that is they are one hop away. If they are more than one hop away the influence is secondary, or tertiary, etc. We assume that all links have the same weight. Consider two players: player one P_1, and player two P_2 that are one hop away. How may P_1 influence P_2, we analyze two difference models of influence.

Two Influence Situations

We will discuss in the next subsection what exactly we mean by "knowledge"[2].

Continuous Influence In this situation knowledge is constantly flowing from one player to another via the link connecting them. What is important to keep in mind is that if knowledge is flowing from n_1 to n_2, the knowledge inherent in n_1 does not diminish. This can be modeled by study of the heat/diffusion equation and the Laplacian and spectral analysis of the graph adjacency matrix. We do not concentrate on this important, but well-studied area of spectral analysis. Rather we concentrate on what follows. Also the flow per link is not instantaneous, it is given in units of knowledge per time per hop.

Probabilistic Influence In this situation there may be resistance to knowledge being passed from n_1 to n_2 via a connecting link. We also assume that we have a clock and that every time unit t, now viewed in a discrete manner, we attempt to pass the knowledge. Thus, the passing of the knowledge is modeled as a geometric random variable with parameter p.

In either situation how does knowledge passed from n_1 to n_2 affect the efficiency?

Knowledge and Changes in Joint Entropy

Let us consider the simple situation of two binary random variables $X \in \{x_1, x_2\}$ with probabilities $p(x_i)$ and $Y \in \{y_1, y_2\}$ with probabilities $p(y_j)$. The joint probabilities are given, as usual, by $P(X = x_i, Y = y_j) = p_{i,j}$. We start off with X and Y being independent, that is $p_{i,j} = p(x_i) \cdot p(y_j)$.

Network Structure and Flow of Knowledge

The topology of the team impacts how knowledge flows, and efficiency improves over time, for a team. Consider a hierarchical team, with a top node node, two nodes under the top node, two nodes under each secondary node, etc. We see that the number of levels of nodes affects the time it takes knowledge to spread through the team. This statement is true whether we consider continuous or probabilistic propagation. However, if we consider a team to be completely connected (in the graph sense), then we see that the spread of knowledge is much faster. Therefore, we must take into account the team topology.

[2] Note that social knowledge is a form of cultural knowledge.

Many models of propagation on a network are given by using epidemiological models. We use this for knowledge transfer.

We are given a connected graph G and assume that a node with a virus at time t can only infect a node that it is a "neighbor", that is it is linked to it by a link (one hop), and this other node will be considered to be infected at time $t + 1$, and the probability of this is β. That is to get a node sick, it most be contiguous to a sick node, and it takes one time interval for this to happen. Furthermore, n_i can become infected at $t+1$ via the infection in a neighbor at t with probability β. If $n_i(t) = 1$, that is n_i is sick at t it can become cured at $t + 1$ with probability δ, this is independent of any neighboring nodes attempting to get it sick — this is a strong modeling assumption. That is, a cure works even if there are more virus attacks coming in.

For us β can model how knowledge moves throughout the team, and δ can model how knowledge flow can be thwarted (this is especially true if it is not a closed system).

The largest eigenvalue of the adjacency matrix $\mathbb{A}$ corresponding to the graph G is $\lambda_{\mathbb{A}_{max}}$ which is also called the spectral radius of $\mathbb{A}_{max}$, written as $\rho(\mathbb{A}_{max})$.

Variations on reinfection is a topic of the various virus models in the literature, see (Prakash 2012). The major result in the literature is (we express as a heuristic)

Heuristic (Wang et al. 2003; Chakrabarti et al. 2008; Prakash et al. 2011; Prakash 2012; Jamakovic et al. 2006; Mieghem, Omic, and Kooij 2009; Mieghem 2011; Newman 2010)
A necessary condition, if there are initially infected nodes, for there not to be an epidemic, and for the infection to probabilistically die out is that

$$\frac{\beta}{\delta} < \frac{1}{\lambda_{\mathbb{A}_{max}}} . \tag{10}$$

The various statements and proofs of the above seem to be lacking, or extremely complex. In future work we will present out proof. We do not do it in this paper for reasons of space limitations and applicability to the tastes of the symposium attendees. We sketch the proof ideas in this paper.

Proof sketch of Heuristic:
PART 1: Iterative Inequality
Let $n_i(t)$ be the condition of node n_i at time t viewed as an indicator function. By this we mean that $n_i(t) = 0$ if n_i is healthy at time t, and if it is sick at time t then $n_i(t) = 1$. We denote the state of health at time t of the entire graph by the column vector $\vec{n}(t)$, whose components are the $n_i(t)$. Thus, the initial state of the graph is $\vec{n}(0)$. We are in a probabilistic situation where $0 \le P\big(n_k(t) = 1\big) \le 1$. We define the vector $\vec{P}_t$ to be the column n vector whose k-th component is $P\big(n_k(t) = 1\big)$. To simplify notation we set $p_{k,t} := P\big(n_k(t) = 1\big)$.

We are interested in $p_{k,t}$ or equivalently $p_{k,t+1}$.

$$\begin{aligned} & p_{k,t+1} \\ &= P\big(n_k(t+1) = 1, n_k(t) = 0\big) + P\big(n_k(t+1) = 1, n_k(t) = 1\big) \\ &= P\big(n_k(t+1) = 1, n_k(t) = 0\big) + P\big(n_k(t+1) = 1 | n_k(t) = 1\big) \cdot p_{k,t} \end{aligned} \tag{11}$$

Define $\mathbb{S} := (1-\delta)\mathbb{I} + \beta\mathbb{A}$, and one can show that to the first order that

$$\vec{P}_{t+1} \lessapprox \mathbb{S} \cdot \vec{P}_t \ . \tag{12}$$

Thus, we have the *approximate* iterative inequality

$$\vec{P}_t \lessapprox \mathbb{S}^t \cdot \vec{P}_0. \tag{13}$$

The use of the matrix $\mathbb{S}$ is not surprising based upon the type of epidemiological virus spread model we are using. Basically we assume that a node maybe healed and not infected in the same time slice. Healing comes from itself, hence the identity matrix $\mathbb{I}$, and infection comes from a neighbor, hence the adjacency matrix. We note that it is not coincidental that the above equations are very similar to the equations given in (Blanchard and Volchenkov 2009, 2.2.1). This is because that book analyzes random walks, which are very similar to what we are doing. A difference is that in the random walks under consideration something either probabilistically stays put or moves to a particular neighbor. In our situation it may move to every neighbor. However, this is a slight difference. In fact, after a (partial) literature review, it seems that (Blanchard and Volchenkov 2009) is the only complete proof of the result.

PART 2: Eigenvalues (This is standard Linear Algebra included for completeness, the interesting modeling work is in part 1.)

Let us compare the spectrums (set of matrix eigenvalues) $spec(\mathbb{A})$ with $spec(\mathbb{S})$. It is trivial to see that

$$\begin{aligned} & \lambda \in spec(\mathbb{A}) \implies 1 - \delta + \beta\lambda \in spec(\mathbb{S}), \\ & \text{and } \tilde{\lambda} \in spec(\mathbb{S}) \implies \frac{\tilde{\lambda} - 1 + \delta}{\beta} \in spec(\mathbb{A}) \ . \end{aligned} \tag{14}$$

Therefore,

$$\rho(\mathbb{S}) = 1 - \delta + \beta\rho(\mathbb{A}) \ , \text{ where } \rho \text{ is the spectral radius of a matrix.} \tag{15}$$

We see, accepting Eq. (13) as an *inequality*, that

$$\lim_{t\to\infty} \mathbb{S}^t = [0] \implies \lim_{t\to\infty} \vec{P}_t = \vec{0}$$

regardless of the initial condition $\vec{P}_0$.

Since $\mathbb{A}$ is symmetric, so is $\mathbb{S}$. The Spectral Theorem (Lang 1971, XIV§13) states that there is an orthonormal basis of $\mathbb{R}^n$ of eigenvectors of $\mathbb{S}$. Thus, if in non-increasing order $spec(\mathbb{S}) = \{\tilde{\lambda}_1 = \tilde{\lambda}_{max} = \rho(\mathbb{S}) > 0, ..., \tilde{\lambda}_n\}$, there is an orthonormal basis $\{\vec{\epsilon}_{\mathbb{S}_1}, ..., \vec{\epsilon}_{\mathbb{S}_n}\}$ then if a generic vector $\vec{v} = \sum_i v_i \vec{\epsilon}_{\mathbb{S}_i}$ then $\mathbb{S}^t \cdot \vec{v} = \sum_i \tilde{\lambda}_i^t v_i \vec{\epsilon}_{\mathbb{S}_i}$. Thus, $\mathbb{S} = \mathbb{U}^{-1} \Lambda\ \mathbb{U}$, where Λ is the diagonal matrix with the eigenvalues of $\mathbb{S}$ down the diagonal, and $\mathbb{U}$ is a unitary matrix.

So, using Perron-Frobenius again, which tells us that $\rho(\mathbb{S}) \geq \mid \tilde{\lambda}_i \mid$ then $\mid \mathbb{S}^t \cdot \vec{v} \mid$ is dominated by $\rho(\mathbb{S})^t \mid v_1 \mid$ which, if $\rho(\mathbb{S}) < 1$, approaches zero as t grows. Therefore,

$$\rho(\mathbb{S}) < 1 \implies \lim_{t\to\infty} \mathbb{S}^t = \lim_{t\to\infty} \mathbb{U}^{-1} \Lambda^t\ \mathbb{U} = [0].$$

In particular, since $0 < \rho(\mathbb{A})$ and using Eq. (15), we have

$$\rho(\mathbb{S}) < 1 \iff \frac{\beta}{\delta} < \frac{1}{\rho(\mathbb{A})} \ . \tag{16}$$

Which is the same as

$$\lambda_{\mathbb{S}_{max}} < 1 \iff \frac{\beta}{\delta} < \frac{1}{\lambda_{\mathbb{A}_{max}}} \ . \tag{17}$$

We note that our approximations and approach were guided by the existing literature (Wang et al. 2003; Chakrabarti et al. 2008; Prakash et al. 2011; Prakash 2012; Jamakovic et al. 2006; Mieghem, Omic, and Kooij 2009; Mieghem 2011; Newman 2010) . However, we note that we only "prove" one way, not if and only if. Because that is not a correct result, the result should be thought of as a phase transition value. Furthermore, we do not find it necessary to appeal to continuous Markov type models, and apply results from the stability of differential equations, for a discrete problem. Thus, we end up with the approximate result:

$$\frac{\beta}{\delta} < \frac{1}{\lambda_{\mathbb{A}_{max}}} \implies \lim_{t\to\infty} \vec{P}_t = \vec{0} \ . \qquad \square \tag{18}$$

Keep in mind that many approximations were done to achieve the above result, and also that it is a probabilistic result. We have run simulations, varying β and δ so that their ratio was constant but we could force an epidemic to live on for a long time or die out quickly *regardless if the above inequality is satisfied*. However, the literature does show that it is a good rule of thumb, and should be considered a phase transition point. This is not surprising because the cutoff value $\frac{1}{\lambda_{\mathbb{A}_{max}}}$ in fact is the boundary between 0 being a locally asymptotic stable fixed point or not (van den Driessche and Watmough 2002).

Table 1: Spectral radius as function of n

n	L_n	R_n	S_n	K_n
2	1	NA	1	1
3	1.412	2	1.412	2
4	1.618	2	1.732	3
5	-	2	-	4
6	-	2	-	5
10	-	2	-	9
n	$2\cos\big(\pi/(n+1)\big)$	2	$\sqrt{n-1}$	n-1
∞	2	2	∞	∞

Discussion

Our heuristic rule of thumb Eq. (10) should really be expressed as an approximation, that is:

$$\frac{\beta}{\delta} \lessapprox \frac{1}{\lambda_{\mathbb{A}_{max}}} \ . \tag{19}$$

We note that as the cure rate approaches zero, the epidemic never dies out — this is not surprising. Even though the above is an approximation if we can lessen $\lambda_{\mathbb{A}_{max}}$ which

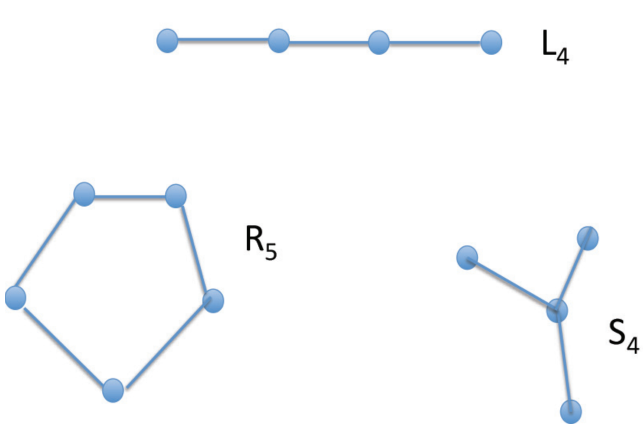

Figure 2: Various Team Topologies

increases $1/\lambda_{\mathbb{A}_{max}}$ we should make it harder for a virus to spread.

We remind the reader that this infection/knowledge transfer model is based upon an exact mathematical model that may not represent all ways for a team to get smarter. However, this model does capture the topological flavor of knowledge transfer in a simple manner that relates to how we may want to model communication paths in a team. Of course, if all players can directly converse with all players (via one hop) we are in the best possible situation.

We must keep in mind though that we cannot just blindly throw mathematical models at the problem. If the knowledge we are discussing is a social or cultural knowledge, then there may exist hard bounds given by the team boundaries.

Spectral radius for various graphs

Let us consider what we have above, keeping in mind that the spectral radius is the inverse of the maximal eigenvalue. Let us state our thinking below.

Spectral Analysis of Team Efficiency The larger the spectral radius of the topology of a team, the more efficient it will be, for all other things being equal.

Let us go back to Table 1. The line graph L_n has the smallest spectral radius. This makes sense, the only way that knowledge can be passed is down the line, and it takes time. If there are 10 people on the team the average number of hops to pass knowledge, and hence influence the dependent random variables is approximately 5. The ring graph R_n is a bit better because we can go to the left of the right, the star graph S_n is even better since we are never more than 2 hops away. Of course the completely connected graph K_n is the best, every player is 1 hop away from another.

Consider Figure 3. We envision a team structure where the manager is the top node, the two assistant managers are on the middle row, and the four workers are on the bottom row. Upon inspection it would appear that ET has better team efficiency as time evolves, than does CT. Lets us look at the maximal eigenvalues of the associated adjacent matrices.

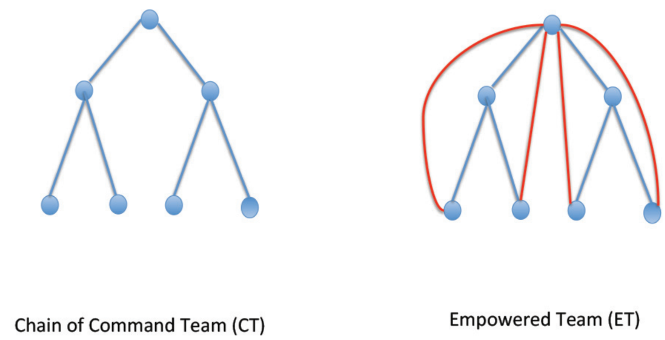

Figure 3: Two Team Structures:top row node 1, middle row nodes 2,3 (left to right), bottom row nodes 4,5,6, & 7 (left to right)

$$\mathbb{CT} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \text{ and}$$

$$\mathbb{ET} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

One can show that $\lambda_{\mathbb{CT}_{max}} = 2$ and $\lambda_{\mathbb{ET}_{max}} = 3.24$, thus our mathematics backs our intuition. Of course, one could obviously look at the team structure and see that in CT, a manager can influence a worker in at best two time steps, where as in ET, it only requires one. We certainly did not need any fancy spectral analysis to see that. Rather, we used this example as a proof of our reasoning.

This reasoning leads us to the following: (Keep in mind that our infection rate is the rate of the spread of knowledge, there is no cure rate. For now at least we do not assume that the players get stupider.)

PARADIGM 1 We make it harder for knowledge to spread by lessening the spectral radius of the connected graph. Mathematically, with connected G and connected G': If $[G] = [G']$, and $\lambda_{G_{max}} < \lambda_{G'_{max}}$, then we say that the team represented by G **performs poorer** than the team T' represented by G'.

PARADIGM 2 Given connected G if we can remove links, but not nodes, and arrive at connected G', so that $[G] = [G']$ and $\lambda_{G'_{max}} < \lambda_{G_{max}}$, then we have **modified** G **into a poorer performer** G', that still connects the same nodes. Of course, turning this around we find ways to increase team performance.

Conclusion

Thus, team topology has a strong affect upon how teams perform. The more connected the are, the better, over time the team should perform. Table 1 shows how the spectral radius (maximal eigenvalue) for various team configurations changes as a function of the number of nodes. L_n is a linear graph, R_n is a ring graph, S_n is a star configuration, and K_n is a completely connected graph (clique).

The new metric we have proposed *Efficiency* is entropy based. The interaction between the random variables modeling how team players behave is influenced by the other random variable, via the various links in the team graph structure. A random variable may not directly influence another via a direct hop, it might take several hops, and if probabilistic effects are incorporated into the modeling the graph theoretic approach applies.

Acknowledgements

We appreciate the assistance of NRL's summer 2014 SEAP students, Indrani Ray and Rebekah Kang. We especially appreciate the time and graciousness of Dr. Stephen Russell of NRL.

References

Balch, T. 1997. Social entropy: a new metric for learning multi-robot teams. In *Proc. 10th International FLAIRS Conference*.

Balch, T. 2000. Hierarchic social entropy: An information theoretic measure of robot group diversity. *Autonomous Robots* 8:209–237.

Blanchard, P., and Volchenkov, D. 2009. *Mathematical Analysis of Urban Networks*. Springer.

Chakrabarti, D.; Wang, Y.; Wang, C.; Leskovec, J.; and Faloutsos, C. 2008. Epidemic thresholds in real networks. *ACM Trans. Inf. Syst. Secur.* 10(4):1:1–1:26.

Conant, R. C., and Ashby, W. R. 1970. Every good regulator of a system must be a model of that system. *Int. J. Systems Science* 1(2):89–97.

Conant, R. C. 1974. Information flow in hierarchical systems. *Int. J. General Systems* 1:9–18.

Conant, R. C. 1976. Laws of information which govern systems. *IEEE Transactions on Systems, Man, and Cybernetics* 6(4):240–255.

Cover, T. M., and Thomas, J. A. 2006. *Elements of Information Theory*. Wiley Interscience.

Feinstein, A. 1958. *Foundations of Information Theory*. McGraw Hill.

Giancoli, D. C. 1980, 1991. *Physics, 3rd Ed.* NJ, USA: Prentic Hall.

Jamakovic, A.; Kooij, R.; Mieghem, P. V.; and van Dam, E. 2006. Robustness of networks against viruses: the role of the spectral radius. In *Proceedings of the 2011 IEEE 11th International Conference on Data Mining*, Symposium on Communications and Vehicular Technology, 35–38. IEEE.

Kenny, D. A.; Kashy, D. A.; and Bolger, N. 1998. Data analysis in social psychology. In Gilbert, D. T.; Fiske, S. T.; and Lindzey, G., eds., *The handbook of social psychology, Vols. 1 and 2 (4th ed.). , (pp. 233-265). New York, NY, US: McGraw-Hill, x, 1085 pp.*

Lang, S. 1971. *Algebra*. Addison Wesley.

Lawless, W. F.; Llinas, J.; Mittu, R.; Sofge, D.; Sibley, C.; Coyne, J.; and Russell, S. 2013. Robust intelligence (RI) under uncertainty: Mathematical and conceptual foundations of autonomous hybrid (human-machine-robot) teams, organizations and systems. *Structure & Dynamics 6(2)*.

Mieghem, P. V.; Omic, J.; and Kooij, R. 2009. Virus spread in networks. *IEEE/ACM Trans. on Networking* 17(1):1–14.

Mieghem, P. V. 2011. The n-intertwined SIS epidemic network model. *Computing* 93:147–169.

Newman, M. 2010. *Networks an Introduction*. Oxford.

Prakash, B. A.; Chakrabarti, D.; Faloutsos, M.; Valler, N.; and Faloutsos, C. 2011. Threshold conditions for arbitrary cascade models on arbitrary networks. In *Proceedings of the 2011 IEEE 11th International Conference on Data Mining*, ICDM '11, 537–546. IEEE Computer Society.

Prakash, B. A. 2012. *Understanding and Managing Propagation on Large Networks—Theory, Algorithms, and Models*. Ph.D. Dissertation, Computer Science Department, Carnegie Mellon University.

Salas, E.; Fiore, S. M.; and Letsky, M. P. 2012. *Theories of Team Cognition*. Routledge, Taylor & Francis Group.

Shannon, C. E. 1948. A mathematical theory of communication. *Bell Systems Technical Journal* 27:379–423,623–656.

Shannon, C. E. 1956. The bandwagon. *IRE Transaction on Information Theory* 2(1):3.

Stanton, N.; Hedge, A.; Brookhuis, K.; Salas, E.; and Hendrick, H. 2004. *Handbook of Human Factors and Ergonomics Methods*. CRC Press.

Surowiecki, J. 2005. *The Wisdom of Crowds*. Anchor Books, first anchor books edition edition.

van den Driessche, P., and Watmough, J. 2002. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences* 180:29–48.

Wang, Y.; Chakrabarti, D.; Wang, C.; and Faloutsos, C. 2003. Epidemic spreading in real networks: An eigenvalue viewpoint. In *In SRDS*, 25–34.