Measuring information integration in Social Networks

D. Calitoiu* and **Zachary Jacobson†**

Abstract

This paper applies the information integration in social networks as a concept. The corresponding measures were proposed by Tononi *et al.*[3, 4] to describe the difference between connectional organizations of certain neural architectures: the thalamocortical systems are well suited to information integration in contrast to the cerebellum, which is not. Our goal is to *adapt* and to *apply* information integration to social networks. The results are novels, and to the best of our knowledges, hitherto in the field of Social Network.

1 Background

The goal of this research is to determine the subnet in a network that has the greatest capacity to integrate information compared with any other subnetworks that can be obtained from the initial network.

To solve this, we used methodologies proposed by Tononi *et. al.* [3, 4, 5, 6] to describe the difference between connectional organization of certain neural architectures, e.g. the thalamocortical systems are well suited to information integration whereas the cerebellum is not. For consistency, a table of definitions and notations used by Tononi is included (see Table 1).

Tononi *et al.* defined a measure, effective information, a quantity that describes all causal interactions that can occur between two parts of a system. They defined Φ the capacity to integrate information, as the minimum amount of effective information that can be exchanged across any possible bipartition of a subset. This measure was used to identify the subsets of a system that can integrate information, called "complexes". Tononi *et al.* applied their analysis to idealized neural

*School of Computer Science, Carleton University, Ottawa, Canada and Health Policy Research Division, Health Canada, E-mail: dcalitoi@scs.carleton.ca.

†Health Policy Research Division, Health Canada, E-mail: zachary_jacobson@hc-sc.gc.ca

X	An isolated system with n elements
S	A subset of X
A and B	A and its complement B : partition in S
CON(S)	Matrix of "anatomical" connectivity
H	Entropy
MI	Mutual Information
EI	Effective Information
MIB	Minimum Information Bipartition
Φ	Capacity to Integrate Information

Table 1: The notations used by Tononi *et al.*.

systems that can have different topologies. Tononi dealt with neural networks, which are digraphs. We apply Tononi's methodologies in Social Networks, considering as networks containing undirected edges.

1.1 Our problem

Consider a network X with n nodes and the edges[1] associated with them $CON(X)$. We will compute for this network, the entropy (H), the mutual information (MI) and the capacity to integrate information, Φ, using Tononi methodologies,

1.2 Information and Entropy

Consider each node of the network X as a source that can generate a message to a different node, if there exists a connection between one node (the source) and the second node (the destination or receptor). Such a source has a certain communication-theory entropy per message, equal to the number of binary digits necessary to transmit a message generated by the source. It takes a specific energy per binary digit to transmit the message against a noise from source to destination [2]. The message reduces the uncertainty of the state of the system, thus decreasing the entropy of the system. The process of reducing the entropy increases the free energy of the system with a value equal to the minimum energy necessary to transmit the message which lead to the increase of free energy, an energy proportional to the entropy of the communication theory [1]. This is, of curse, the relation between the *entropy of communication theory* (the entropy related to the transmission process) and *the entropy of statistical mechanics* (the entropy related to source of transmission).

[1]The term CON is obtained from the abbreviation of *connections*, the links between neurons.

Pierce stated [1] that one pays a price for information which leads to a reduction of the statistical-mechanical entropy of a system and the price is proportional to the communication-theory entropy of the message source which produces the information.

1.3 Effective information

Consider S a subnetwork of the initial network X, that contains n_S nodes, namely $n_S < n$. It is possible to measure the information generated by the network S when it is in an arbitrary state, namely we know all the values associated with the nodes and with the edges. Consider a repertoire of all the possible value that the subnetwork (with n_S nodes and the corresponding edges described by $CON(S)$) can have. We now generate an arbitrary partition, dividing the network in two subnetworks, A and its complement B, namely $B = S - A$. We denote this partition of the network S in two disjoints subnetworks as $[A : B]_S$. We consider the case where A has the maximum entropy $A^{H^{max}}$[2]. We evaluate effect in B if A has this setting, namely the maximum entropy. In this context, it is possible to define the effective information from A to B as :

$$EI(A \to B) = MI(A^{H^{max}} : B) \tag{1}$$

where

$$MI(A : B) = H(A) + H(B) - H(AB) \tag{2}$$

is the mutual information, namely the measure of the entropy or information shared between a source A and a target B. One remark can be done related to the value of $EI(A \to B)$. If the edges that connect A with B are strong, $EI(A \to B)$ will he high. When the edges that connect A with B are weak, the value of $EI(A \to B)$ will be low. The effective information measures the causal interaction between A and B and it is mandatory in order to do the computation to insert an information in A and after that to evaluate the effect in B. This approach is different from the measures that describe the statistical dependence between A and B. In this respect, note that in general $EI(A \to B)$ is different than $EI(B \to A)$, thus the effective information is not symmetric. For any arbitrary disjoint partition $[A : B]_S$, the effective information is the sum of the effective

[2]We will explain latter how is possible to compute the maximum entropy associated with a network.

information for both directions:

$$EI(A \rightleftharpoons B) = EI(A \to B) + EI(A \leftarrow B) \tag{3}$$

1.4 Information integration

Defining the effective information for any arbitrary disjoint partitions $[A:B]_S$, forces the question of how much information can be integrated in a subnetwork. The level of integration depends on the effective information, because in the case where $EI(A \rightleftharpoons B) = 0$ the network can not integrate any information: A and B are two causally independent subnetworks. Tononi proposed to measure how much information can be integrated within a network S, by searching for the bipartition $[A:B]_S$ for which $EI(A \rightleftharpoons B)$ reaches a minimum. This minimum is generated by the existence of the weakest link in the subset: there can only be as much integrated information within the subset as there is across its minimum cut. The effective information computed for this arbitrary disjoint partition $[A:B]_S$, namely $EI(A \rightleftharpoons B)$, has the property that is bounded by the smallest between the maximum entropy available to A or B. In order to be comparable over partitions, the values $min\{EI(A \rightleftharpoons B)\}$ is normalized by $min\{H^{max}(A), H^{max}(B)\}$. After this normalization, the *minimum information bipartition* of subset S, or $MIB(S)$, is its bipartition for which the normalized effective information reaches a minimum, that is evaluated as:

$$MIB(X) = [A:B]_S \tag{4}$$

where

$\min\{EI(A \rightleftharpoons B)\}/(\min H^{max}(A), \min H^{max}(B)) = \min$ for all A.

With these concepts introduced, Tononi proposed the *capacity for information integration* of subsets X, or $\Phi(X)$[3] , that is the value of $EI(A \rightleftharpoons B)$ for the minimum information bipartition:

$$\Phi(S) = EI(MIB(S)). \tag{5}$$

This quantity is also called $MIB^{complexity}$ for minimum information bipartition complexity.

[3]Tononi mentioned that the Greek letter Φ is used to indicate the information (the "I" in "Φ") that can be integrated within a single entity (the "O" in "Φ").

Example 1 from Tononi	Example 2 from Tononi	The Sampson Monastery network
73.6039 bits	20.5203 bits	85.292395 bits

Table 2: Evaluation of Φ for three networks.

1.5 Complexes

Consider again the premises of our initial problem. We have a network X with n nodes and the edges associated. It is possible to select an arbitrary network S from this network, that that its number of nodes $n_s < n$. Consider now that it is possible to generate *all* the subsets with n_s, starting with $n_s = 2$ and ending with the subset corresponding to he entire[4] networks $n_2 - n$. Each subset has a corresponding $\Phi(S)$ value. For a given network X, the subnetwork S with the maximum value of $\Phi(S)$ is called the main complex, where the maximum is taken over all combinations of $n_s \geq 2$ out of n nodes of the network:

main complex(X)= $S \subseteq X$ for which $\Phi(S) = \max$ for all S.

2 Results

We applied the Φ measure to different networks. Unfortunately, the procedure of computing Φ, presented in Annex 1, involves a *great* computational power. With a ordinary computer (Pentium 4, 2.8 Ghz processor, and 1 Gb RAM), for a networks with 18 nodes, around 6 days are required for computations. In this paper we present the results for only one dataset[5] Sampson Monastery (a network with 18 nodes). The details of this dataset are presented in Annex 2. The network is depicted in Figure 2, using NetDraw[6].

We compared our network with the ones generated by Tononi in [3], namely two networks with only 8 nodes (See Figure 1). In all the computations we used the following settings: w=0.5 (the weight of the connections), $c_i = 0.00001$ and $c_p = 1$. (See Annexa 1 for details).

The comparison between the computed Φ for these three networks can be seen in Table 1.

The results prove the influence of homogeneity and modularity of connection pattern on the integration of information. The value of Φ corresponding to the Example 1, a network with 8 nodes,

[4]This extreme case when $n_2 - n$ is without any relevant value when we are investigating real networks with relatively large number of nodes.

[5]The dataset used in this paper are provided with UCINET6 that can be downloaded at www.analytictech.com.

[6]Netdraw is a free social network visualization package and can be downloaded at www.analytictech.com.

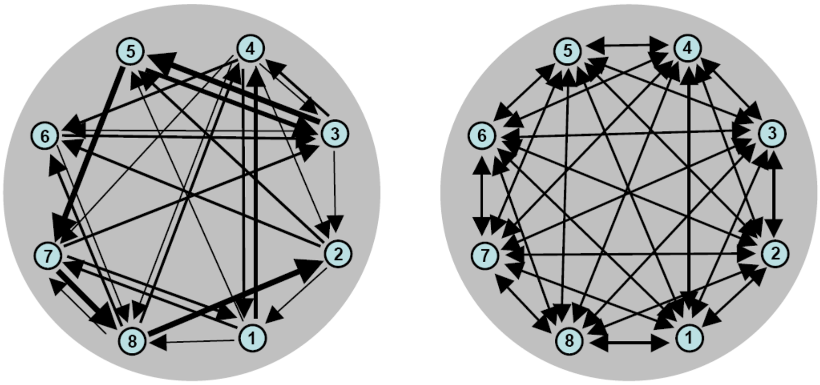

Figure 1: Two examples, Example 1 (Left) and Example 2 (Right) taken from Tononi [3]. The networks contain 8 nodes. For Example 1, there is an heterogeneous arrangement of the incoming and outgoing connections for each element. For Example 2, connectivity is full.

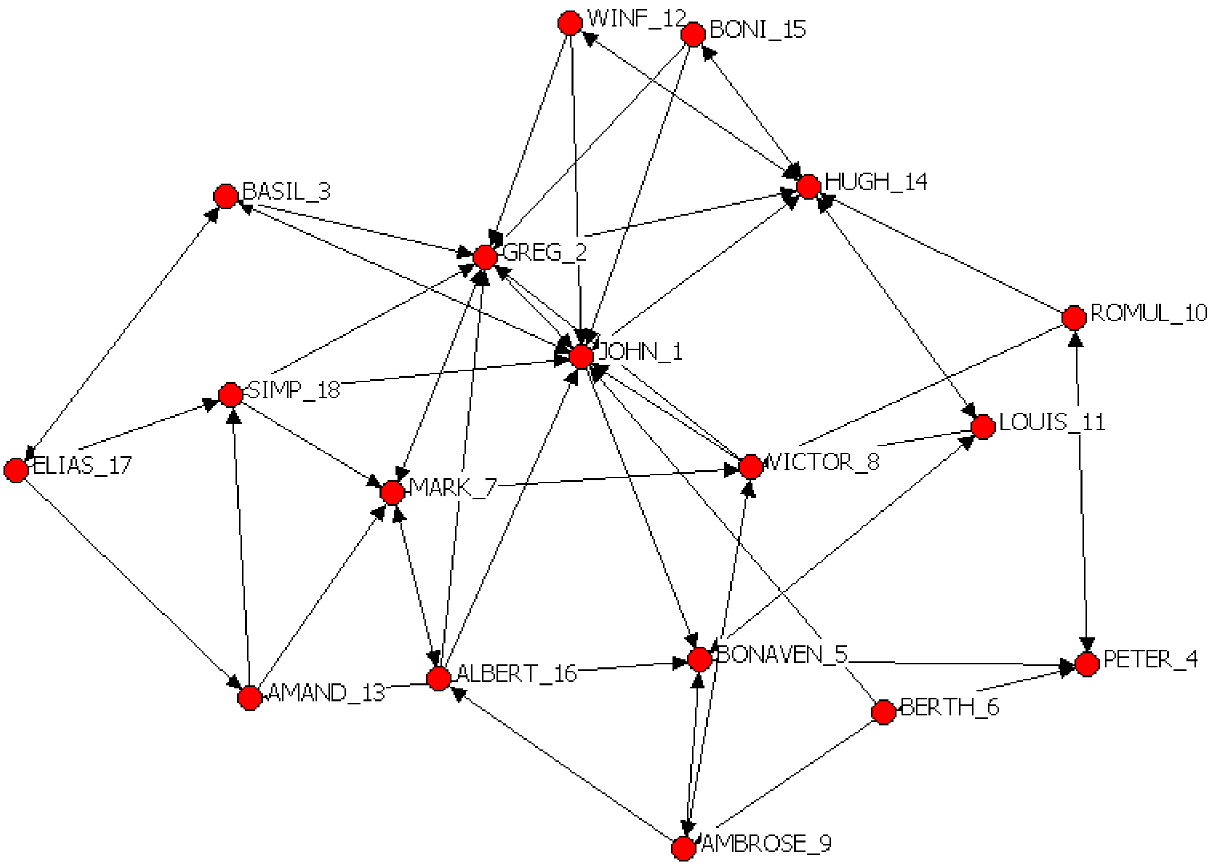

Figure 2: Sampson Monastery network (18 nodes)

is comparable with the value of Φ corresponding to Sampson Monastery network that contains 18 nodes. The connections added to form the homogenous network presented in Example 2 have the effect of decreasing the capacity of integrating information in the network.

References

[1] J.R. Pierce. An Introduction to Information Theory (Second edition). *Dover Publications. Inc. New York*, 19808.

[2] C.E. Shanon. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27:623-656, 1948.

[3] G. Tononi and O. Sporns. Measuring information integration. *BMC Neuroscience*, 4:31, 2003.

[4] G. Tononi. An information integration theory of consciousness. *BMC Neuroscience*, 5:42, 2004.

[5] G. Tononi and G.M. Edelman Consciousness and complexity. *Science*, 282(5395):1846-1851, 1998.

[6] G. Tononi and O. Sporns and G.M. Edelman A complexity n information integration theory of consciousness. *BMC Neuroscience*, 5:42, 2004.

Annex 1: The method of computing Φ

The algorithm of computing Φ is adapted from Tononi *et al.* [3].

1. Computing the normalized matrix related to connections

   Consider a network with n nodes and the corresponding edges that link these nodes stored in the matrix CON_{ij}. The self-connection are excluded. The CON_{ij} matrix is normalized with the following rule: the absolute value of the sum of the afferent synaptic weights per node correspond to a constant value $w < 1$. The new normalized matrix is denoted $CON(X)$.

   We consider a vector $\mathbf{X}$ of random variables that represents the activity of the nodes of the network, that consist in a independent Gaussian noise $\mathbf{R}$ of magnitude c. This specific type of random variable is chosen for allowing a maximum value for the entropy generated from each nodes. The details will follow.

When the elements are configured under stationary conditions,

$$\mathbf{X} = \mathbf{X} * CON(X) + c\mathbf{R}. \tag{6}$$

2. Computing the covariance matrix related to connections

   We define $Q = (1 - CON(X)))^{-1}$ and we can average over the states produced by successive values of $\mathbf{R}$. In this way, a covariance matrix is computed:

$$COV(X) = \langle \mathbf{X} * \mathbf{X} \rangle = \langle Q^t * \mathbf{R}^t * \mathbf{R} * Q \rangle = \langle Q^t * Q \rangle. \tag{7}$$

   In the above equation, the superscript t is used to describe the transpose of the matrix.

3. Computing the entropies and the mutual information

   Considering the Gaussian assumption, all the deviation from independence among the two disjoints parts A and B of the subnetwork S of a network X are described by the covariances among the respective nodes. Knowing these covariances, one can compute the individual entropies $H(A)$ and $H(B)$ and the corresponding joint entropy of the subset $H(S) = H(AB)$, using the formula $H(A) = \frac{1}{2}\ln[(2\pi e)^n \det\{COV(A)\}]$ where $\det\{Y\}$ denotes the determinant of Y. With these entropies computed one can obtain the mutual information between the partitions A and B, namely $MI(A : B) = H(A) + H(B) - H(AB)$.

4. Computing the effective information

   For computing the effective information between subnetworks A and B, one must enforce independent noise sources in the nodes of network A by setting to zero strength the connections within A and afferent to A. In this configuration, we have the followings: (i) the covariance matrix that corresponds to the network A is equal to the identity matrix (under independent Gaussian noise assumption) and (ii) the statistical dependence between A and B is based on the causal effects of subnetwork A on subnetwork B, mediated by the efferent connections of subnetwork A. Additionally, all the possible outputs form the nodes of subnetwork A that could affect the subnetwork B are evaluated. With this rationale, to finalize the computation of $EI(A \rightarrow B) = MI(A^{Hmax} : B)$ there are two parameters introduced:

(i) the independent Gaussian noise $\mathbf{R}$ applied to each node of the subnetwork A is multiplied by the *perturbation coefficient* c_p;

(ii) the independent Gaussian noise applied to the rest of the network is is multiplied by the *intrinsic noise coefficient* c_i.

The values associated with c_p and c_i can modify the signal-to-noise (SNR) of the effects of subnetwork A on the rest of the network. With $c_p = 1$ and $c_i = 0.00001$, the SNR is high; in contrast, with $c_p = 1$ and $c_i = 0.1$, the SNR is low. In the scenario when the connections within and to the network A are NOT modified and the independent noise sources are associated with the nodes of the subnetwork A, one would obtain a modified effective information (and derived measures) that reflect the probability distribution that corresponds to the subnetwork A as filtered by its connectivity.

5. Computing the minimum information bipartition and the capacity for information integration

   Consider every subset $S \subseteq X$ containing $K = 2, ..., n$ nodes. For each disjoints partitions A and B compute $EI(A \rightleftharpoons)$ and find (i) the minimum information bipartition $MIB(S)$ for which the normalized effective information reaches a minimum and (ii) the corresponding value of Φ. Thus, we find the *complexes* of the networks X as those subnetworks S having $\Phi > 0$ that are not included within a subnetwork having a higher Φ and rank them based on their value of $\Phi(S)$. To conclude, the complex with the maximum value of $\Phi(S)$ is the *main complex*, as defined by Tononi *et al.* [3].

Annex 2 : The dataset

2.1 Sampson Monastery (18 nodes)

Background: Sampson recorded the social interactions among a group of monks while resident as an experimenter on vision, and collected numerous sociometric rankings. The labels on the data have the abbreviated names followed by the codings used by Breiger and Boorman in all their work (See Figure 2). During their presence in the monastery, a political “crisis in the cloister” resulted in the expulsion of four monks (No. 2, 3, 17, and 18) and the voluntary departure of several others - most immediately, No. 1, 7, 14, 15, and 16. (In the end, only 5, 6, 9, and 11 remained). All the numbers used refer to the Boorman and Breiger numbering. Hence in the end Bonaventure, Berthold, Ambrose and Louis all remained.

Most of the present data are retrospective, collected after the breakup occurred. They concern a period during which a new cohort entered the monastery near the end of the study but before the major conflict began.

References:

1. Breiger R., Boorman S. and Arabie P. (1975). *An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling.* Journal of Mathematical Psychology, 12, 328-383.

2. Sampson, S. (1969). *Crisis in a cloister.* Unpublished doctoral dissertation, Cornell University.