Entropy rates of spatially distributed state machines

James Rising

Information theory has much to contribute to the study of cellular automata. This paper considers several consequences for "binary conformist grids", cellular automata networks where each node aims to have a state similar to its neighbors. These results have consequences for the movement of information and memes through social networks.

A cellular automata consists of a network of nodes, each of which has a state which changes in time. The state evolves according to a state machine, informed by both values within the node's neighborhood.

Throughout this paper, I will use the following notation. The random variable describing the state of a given node at a given point in time is X or X_0. The state of the node's N neighbors is X^n. Often, only number of neighbors of value 1 will matter, and this is denoted by X_n. The state of the remainder of the graph with the exception of node i is X^{-i} and the remainder without node i or its neighbors is $X^{-n_i - i}$. The states of the same nodes at the next moment in time is Y. (e.g., Y, Y^n, etc.).

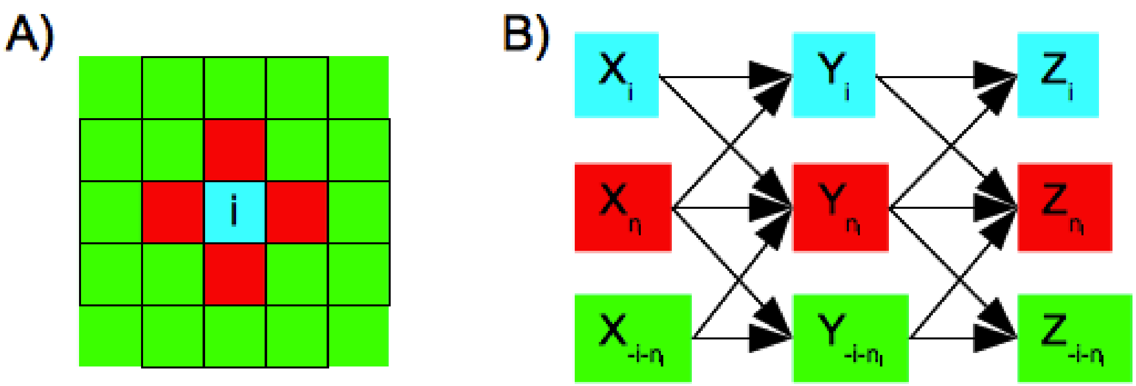


Figure 1: A) Regions of a conformist grid, relative to a node i. Red represents the neighborhood, green represents the remaining grid. B) Markov chain representation of the evolution of states for the regions in (A).

The state machines studied in this paper are non-deterministic, where the evolution of a node's state is a draw from the probability distribution $P(Y|X^n, X)$.

In a binary conformist grid, the more values within a node's neighborhood which are 1, the more likely the node's subsequent state will be 1. This function, $P(Y = 1|X^n, X) = g(X^n, X)$ can take several interesting forms. Below I consider example cases where $q(\cdot)$ is (1) linear or (2) follows the portion of 1's with limits, and where the significance of neighbors are weighted equally or not.

1 Omni-Conformist Grids

Consider a simple state machine, corresponding to a omni-conformist network of size 1. The probability of maintaining the state over time $1 - p$.

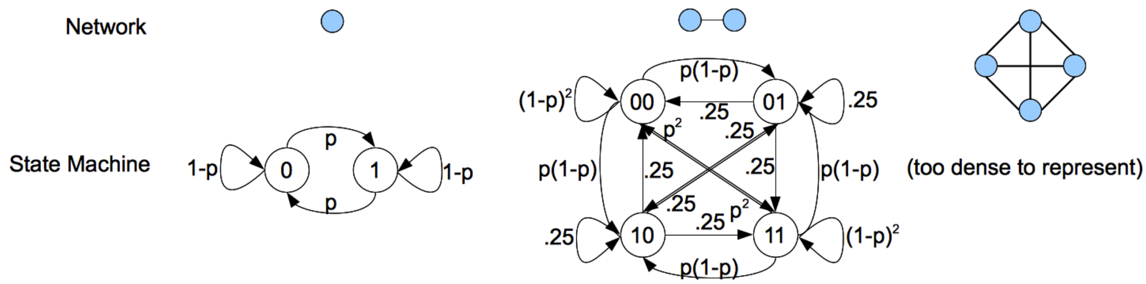


Figure 2: Omni-Conformist Networks, of size 1, 2, and 4.

This can be represented as a Markov chain, with

$$P = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} \qquad \bar{\mu} = \begin{pmatrix} 1/2 & 1/2 \end{pmatrix}$$

Such that $\bar{\mu} = \bar{\mu} P$. For this network,

$$\begin{aligned} \mathcal{H}(\mathcal{X}) &= -\sum_{i,j} \bar{\mu}_i P_{ij} \log P_{ij} \\ &= H(p) = -p \log(p) - (1-p) \log(1-p) \end{aligned}$$

Now, consider an omni-conformist network of size N. There are 2^N states in the composite state machine, each characterized by the set of node states represented by N binary digits. The general form for the transition matrix is,

$$P_{ij} = (1 - \bar{g}(\mathbf{1}(i)))^{n - \mathbf{1}(j)} \bar{g}(\mathbf{1}(i))^{\mathbf{1}(j)}$$

where $\bar{g}(k)$ is the probability that any given bit will be 1 after observing k 1's, and $\mathbf{1}(j)$ is the number of 1's in state j. In my examples, I consider two forms for $\bar{g}(k)$:

$$\begin{aligned} \bar{g}_1(k) &= p + (1-2p)\frac{k}{n} \\ \bar{g}_2(k) &= \begin{cases} p & \text{if } k = 0 \\ 1-p & \text{if } k = n \\ \frac{k}{n} & \text{else} \end{cases} \end{aligned}$$

Elsewhere in this paper, I generally use $g(x^n) = \bar{g}(\mathbf{1}(x^n))$, where now $\mathbf{1}(\cdot)$ denotes the number of 1's in x^n.

Let the steady-state be denoted $\bar{\mu}^N$. It is solveable, but has no closed-form solution for an arbitrary number of nodes, under either of my example forms for $\bar{g}$.

The transition from state i has an entropy $H(B(n, \bar{g}(i))) = nH(\bar{g}(i))$. That is, it has the entropy of a draw from a Binomial distribution, which has n times the entropy of a draw from a Bernoulli distribution. So,

$$\mathcal{H}(\mathcal{X}) = n \sum_{i=1}^{N+1} \bar{\mu}_i H(\bar{g}(i-1))$$

Since each node only observes the portion of nodes in each state, the total number of effective states can be reduced to $n+1$. In this case, the general form of the transition matrix is

$$P_{ij} = \binom{n}{j-1} (1 - \bar{g}(i))^{n-j+1} \bar{g}(i)^{j-1}$$

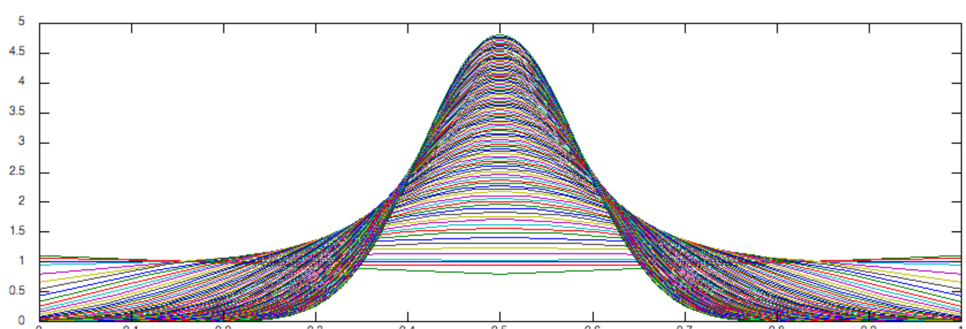
If only the "effective state" is observed– that is, only the count of nodes in each state– the entropy rate decreases significantly. In particular, the relationship between the entropy rates is,

$$\bar{\mathcal{H}}(\mathcal{X}) = \mathcal{H}(\mathcal{X}) - \sum_i \bar{\mu}_i \log \binom{n}{i-1}$$

See figure 3 for a collection of the PDFs.

1.1 Shapes of the Equilibrium

The grid-wide distribution is greatly affected by the impact of minority values. The graphs in figure 3 use $\bar{g}_2(k)$ above. However, when $\bar{g}_1(k)$ is used, the resulting equilibria are very different.



While $\bar{g}_2(k)$ maintains the portion of values of a particular minority, $\bar{g}_1(k)$ increases them slightly.

The effects are clearly visible for an $N = 5$ network. Here, the difference between $\bar{g}_1(k)$ and $\bar{g}_2(k)$ is only for $k = 1$ and $k = 4$. In the graph below, I replace these values with a value p which varies from .25 to .35, and graph the resulting steady-state, $\bar{\mu}$.

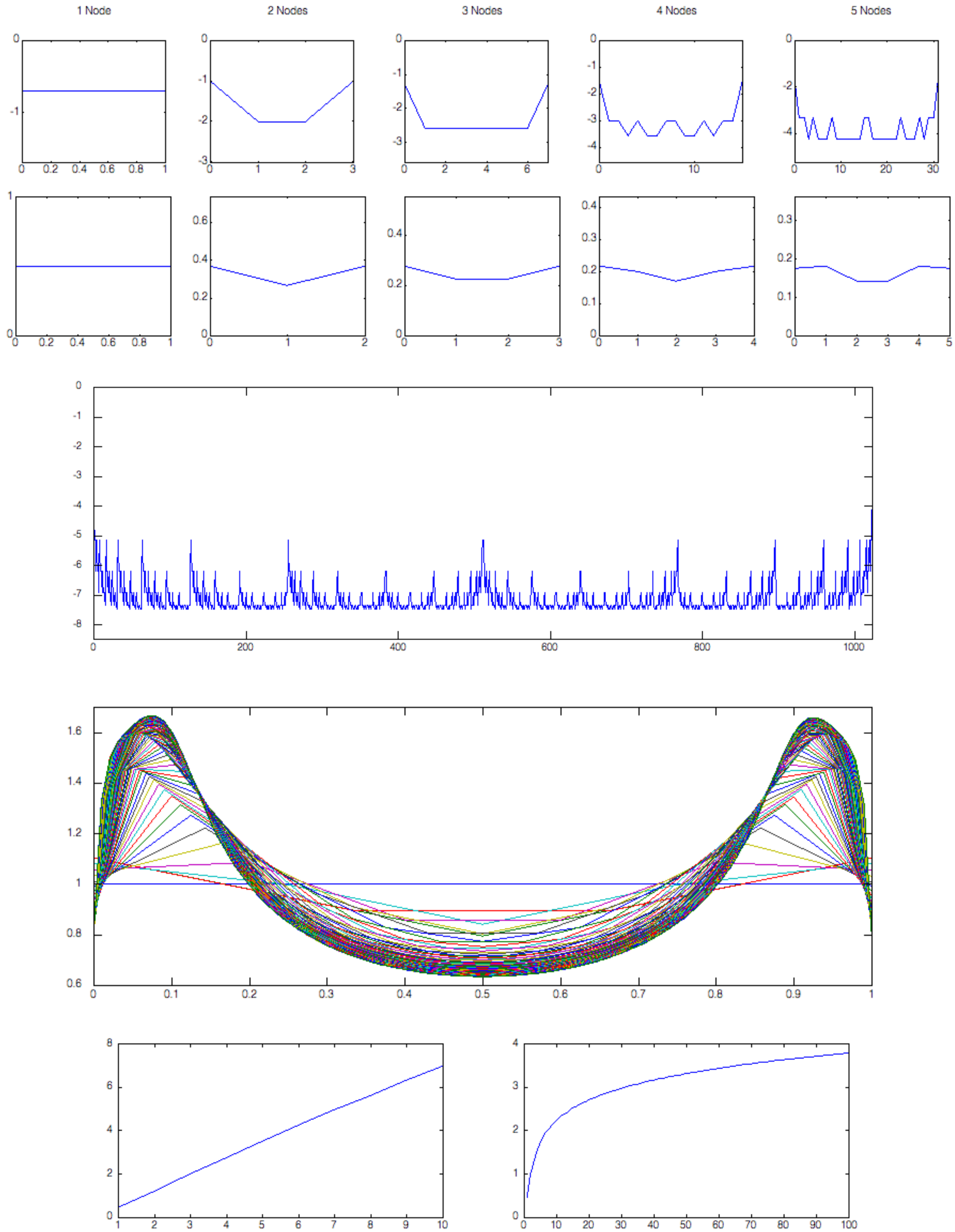
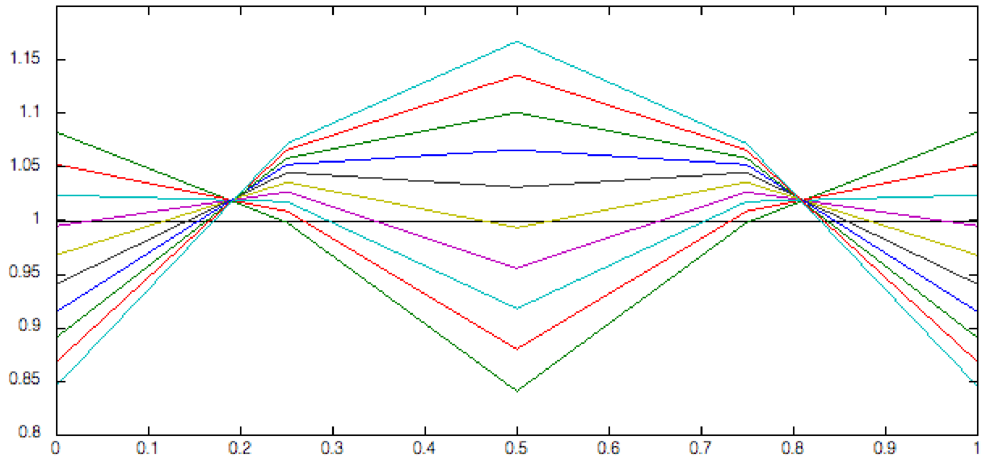


Figure 3: Examples, with $g(i)$ defined as above. Top: PDFs for equilibrium states, by node and in aggregate. Next: PDF for 10 node equilibrium state. Next: PDFs for all equilibrium states in the aggregate model of between 1 and 100 nodes. The PDFs are rescaled to have an approximate area of 1, as though they are continuous PDFs for display purposes. Bottom: Entropy rates for observing the complete state or observing the aggregate state.



2 Conformist Grids

Within a conformist grid, each node has a collection of neighbors. In each time step, each node acts like it is part of an omni-conformist network composed of itself and its neighbors.

The grid can be concieved as a Markov chain in space-time.

$$H(X_i|X_{-i}) = H(X_i|X_{n_i})$$

2.1 Weighted Neighbors

The equations below use a general function $g(x^n)$, as above. This can be modified to allow nodes to apply different weights to different members of its neighborhood. For example, the weighting scheme might count the value of the node itself with higher weight than any of its neighbors, or it might count neighbors further from it less heavily.

The only change needed is in $\mathbf{1}(\cdot)$. Previously, $\mathbf{1}$ was just a function of the number of 1's. Now, we define it as,

$$\mathbf{1}(x_1, \ldots, x_K) = \frac{\sum_k w_k x_k}{\sum_k w_k}$$

2.2 Equilibrium Properties

For an torus grid of size $> N^2$, without edge effects, the equilibrium distribution of states within a region consisting of a node and its neighbors is identical to the omni-conformist network of size $N + 1$.

Proof: At a given time, let the distribution of every node and its associated neighbors be $\bar{\mu}^{N+1}$.

According to the time evolution, for each possible state i, each node draws from a Bernoulli distribution with $p = \bar{g}(i - 1)$. Given the distribution of states around them, they draw independently, so the resulting distribution of states is $\bar{\mu}^{N+1}$, so this is the steady-state.

2.3 Adjacent Information

Under equilibrium,

$$
\begin{aligned}
H(X^n|X) &= H(X, X^n) - H(X) \\
&= H(\bar{\mu}^{N+1}) - 1
\end{aligned}
$$

$$
\begin{aligned}
& H(X, X_1, X_{n-1}) = H(X, X_1) + H(X_{n-1}|X, X_1) \\
\Longrightarrow\ & H(X, X_1) = H(X, X_1, X_{n-1}) - H(X_{n-1}|X, X_1) \\
\Longrightarrow\ & H(X, X_1) = H(\bar{\mu}^{N+1}) - H(\bar{\mu}^{N+1}|X, X_1) \\
& H(X, X_1) = H(X) + H(X_1|X) \\
\Longrightarrow\ & H(X_1|X) = H(X, X_1) - H(X) \\
\Longrightarrow\ & H(X_1|X) = H(\bar{\mu}^{N+1}) - H(\bar{\mu}^{N+1}|X, X_1) - 1 \\
& I(X_1, X) = H(X_1) - H(X_1|X) \\
\Longrightarrow\ & I(X_1, X) = 1 - \left[H(\bar{\mu}^{N+1}) - H(\bar{\mu}^{N+1}|X, X_1) - 1\right] \\
\Longrightarrow\ & I(X_1, X) = 2 - H(\bar{\mu}^{N+1}) + H(\bar{\mu}^{N+1}|X, X_1)
\end{aligned}
$$

Let $\tilde{Y}^n$ be the fictional state of the neighborhood, resulting from an omni-conformist evolution of X, X^n.

$$
\begin{aligned}
& H(\tilde{Y}^n|Y, X) = H(Y, \tilde{Y}^n|Y, X) = H((\bar{\mu}^{N+1}|X)P|Y) \\
& H(Y, \tilde{Y}^n|X) = H(Y|X) + H(\tilde{Y}^n|Y, X) \\
\Longrightarrow\ & H(Y|X) = H(Y, \tilde{Y}^n|X) - H(\tilde{Y}^n|Y, X) \\
\Longrightarrow\ & H(Y|X) = H((\bar{\mu}^{N+1}|X)P) - H((\bar{\mu}^{N+1}|X)P|Y) \\
\Longrightarrow\ & H(Y|X) = I((\bar{\mu}^{N+1}|X)P; Y)
\end{aligned}
$$

3 Non-Equilibrium Properties

In non-equilibrium settings, the probability mass of X^n is not well defined, so the question becomes, if we know the neighborhood of a node, how thoroughly does that constrain the node. If the value of the node and its neighborhood are known at time t, then the entropy increases at time $t+1$ according to

$$
H(Y|X = x, X^n = x^n) = H(g(x, x^n))
$$

Given both x and x^n, y is a draw from the $g(\cdot)$ distribution.

The key non-equilibrium factor is $H(Y|X^n = x^n)$.

Assuming that we have no information on X, let $P(X = 1) = \frac{1}{2}$. Then,

$$P(Y = y|X^n = x^n) = \frac{1}{2}P(Y = y|X = 0, X^n = x^n) + \frac{1}{2}P(Y = y|X = 1, X^n = x^n)$$

$$\begin{aligned} P(Y = 1|X^n = x^n) &= \frac{1}{2}\left[g(0, x^n) + g(1, x^n)\right] \\ &= \frac{1}{2}\left[\bar{g}(\mathbf{1}(x^n)) + \bar{g}(\mathbf{1}(x^n) + 1)\right] \\ P(Y = 0|X^n = x^n) &= \frac{1}{2}\left[1 - g(0, x^n) + 1 - g(1, x^n)\right] \end{aligned}$$

The entropy introduced by time then is,

$$\begin{aligned} H(Y|X^n = x^n) = &- \sum_{y=\{0,1\}} p(y|x^n) \log p(y|x^n) \\ &= -\frac{1}{2}\left[1 - g(0, x^n) + 1 - g(1, x^n)\right] \log \frac{1}{2}\left[1 - g(0, x^n) + 1 - g(1, x^n)\right] \\ &\quad - \frac{1}{2}\left[g(0, x^n) + g(1, x^n)\right] \log \frac{1}{2}\left[g(0, x^n) + g(1, x^n)\right] \end{aligned}$$

The final entropy is

$$H(Y|X^n) = \sum_{x^n \in \mathcal{X}^n} p(x^n) H(Y|X^n = x^n)$$

4 Conformist Grid Channels

Communication can occur over the grids, by designating sets of nodes as transmitters and recievers. Depending on the persepective and grid setup, this can be concieved of as a simple channel, a relay channel, an interference channel, a 2-way (feedback) channel, or a multiple access channel.

4.1 Two-Cell Grids

Consider a grid composed of two cells, a transmitter and a receiver. The receiver recieves the transmission one time step after it is sent. Let the value at the reciever at the time of transmission be X, and the received transmission (the value at the same cell one time step later) be Y. The value being trasmitted is T.

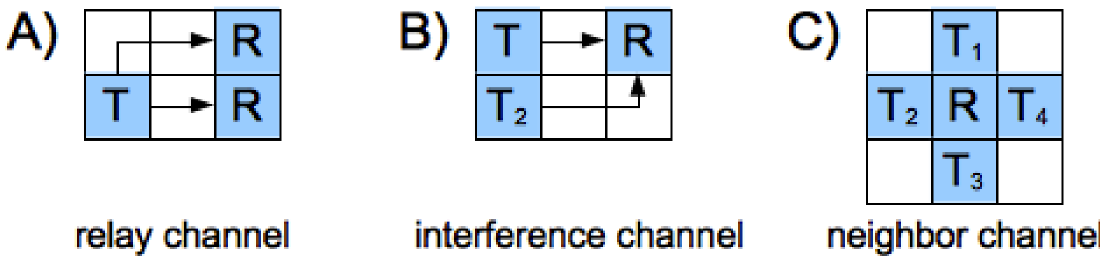


Figure 4: The conformist grid concieved as (A) a relay channel, with multiple recievers to capture the relayed message, (B) an interference channel, with independent transmitters, and (C) the simplest multiple-access channel, the neighbor channel.

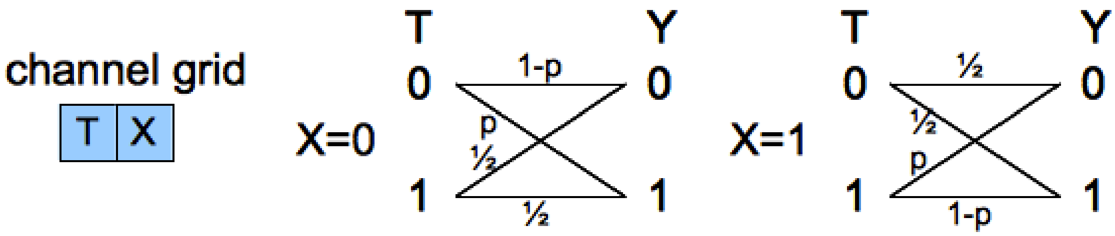


Figure 5: The two channel grid. The entropy depends on whether the value transmitted is the same as the previous value in cell X.

What is the channel capacity, if the channel has memory (it knows the previous value, X)? Let that $q = P(T = 1)$.

$$
\begin{aligned}
C = \max_{p(T)} I(T, Y|X) &= H(T) - H(T|Y, X) \\
&= H(q) - \sum_{x \in \{0,1\}} p(X = x) H(T|Y, X = x)
\end{aligned}
$$

Since $H(T|Y, X = 0) = H(T|Y, X = 1)$, without loss of generality, I will let $X = 0$.

$$P(T=0, Y=0|X=0) = (1-p)(1-q)$$
$$P(T=1, Y=0|X=0) = \frac{1-q}{2}$$
$$P(T=0, Y=1|X=0) = qp$$
$$P(T=1, Y=1|X=0) = \frac{q}{2}$$
$$P(T=0|Y=0, X=0) = \frac{(1-p)(1-q)}{(1-p)(1-q) + \frac{1-q}{2}}$$
$$P(T=1|Y=0, X=0) = \frac{\frac{1-q}{2}}{(1-p)(1-q) + \frac{1-q}{2}}$$
$$P(T=0|Y=1, X=0) = \frac{qp}{qp + \frac{q}{2}}$$
$$P(T=1|Y=1, X=0) = \frac{\frac{q}{2}}{qp + \frac{q}{2}}$$

Based on these, we can calculate $H(T|Y, X=0)$:

$$H(T|Y, X=0) = -\sum_{t\in\{0,1\}} \sum_{y\in\{0,1\}} p(t, y|0) \log p(t|y, 0)$$

4.2 Full Grid Channels

The simplest communication system on the conformist grid is a multiple-access channel formed by the cells neighbors. In general, the capacity region of this network is the closure of the convex hull of rate vectors satisfying

$$\sum_{i\in S} R_i \leq I(X(S); Y|X(S^c)) \, \forall \, S \subseteq \{1, 2, \dots, N\}$$

with $X(S) = \{X_i : i \in S\}$, for some distribution $\prod_{i=1}^{N} p_i(x_i)$ (Cover and Thomas 2006, Thm 15.3.6). The constraining equation in this case is

$$\sum_{i=1}^{N} R_i \leq I(X_1, \dots, X_N; Y)$$

We can maximize the information with a $P(X_n = 0) = 1$ and $P(X_n \neq 0) = 0$. Let $p = \bar{g}(0)$ and $q = \bar{g}(1)$.

$$
\begin{aligned}
P(Y=0|X_n=0) &= \frac{1}{2}\,(p+q) \\
P(Y=1|X_n=0) &= 1-\frac{1}{2}\,(p+q) \\
H(Y|X_n=0) &= H(\frac{p+q}{2}) \\
I(X_n;Y) &= H(Y)-H(Y|X_n) \\
&= 1-H(\frac{p+q}{2})
\end{aligned}
$$

The same capacity applies to a channel formed by a node which transmits to a receiver collectively formed by its neighbors. This is because, $I(Y^n, X) = I(X^n, Y)$.

This provides an upper bound on the mutual information between a node and its neighbor. The mutual information between two nodes separated by a larger distance, K, is a succession of "information losses" like this. So, the probability that information will be lost between a node and its neighbor is

$$P_1 = 1 - \frac{p+q}{2}$$

The chance that information after $K-1$ steps will be lost on the Kth step is

$$P_K = 1 - P_{K-1}\frac{p+q}{2}$$

And the total loss is,

$$P = \prod_{i=1}^{K} P_i$$

The mutual information then is,

$$
\begin{aligned}
I(Z_K, X) &= H(Z_K) - H(Z_K|X) \\
&= 1 - H(P)
\end{aligned}
$$