Finite Time Consensus of Averaging Processes in Connected Graphs

Warren Jin

March 2024

1 Introduction

Consider the following problem:

Problem 4. Consider a 2018×2019 board with integers in each unit square. Two unit squares are said to be neighbours if they share a common edge. In each turn, you choose some unit squares. Then for each chosen unit square the average of all its neighbours is calculated. Finally, after these calculations are done, the number in each chosen unit square is replaced by the corresponding average. Is it always possible to make the numbers in all squares become the same after finitely many turns?

Figure 1: APMO 2019 Problem 4.

It turns out that the answer to this question is no. By analyzing an equivalent process over the integers mod p, it can be shown that some initial configurations are invariant under the averaging process. This result leads to two questions:

1. Given a graph with starting values defined over its nodes, when can we be guaranteed to have a sequence of "averaging" that makes all the numbers the same after finitely many turns, regardless of the starting values?

2. For a given graph, how many starting values must be deliberately chosen such that there exists a finite sequence of averaging that makes all the numbers the same?

I will attempt to answer the first question, and briefly present a hypothesis on the second.

2 Problem Statement

Let the finite connected graph G have nodes labeled $\{1, ..., n\}$. Let $A(G)$ be the nxn adjacency matrix of G. Let $\vec{g}$ be a nx1 vector representing the initial values assigned to the respective nodes in G, such that the first value in $\vec{g}$ is assigned to the node labeled 1, and so on. WLOG we can assume that $||\vec{g}|| = 1$, or else we can simply take $\hat{g} = \frac{\vec{g}}{||\vec{g}||}$ to be the original values.

Define P_i to be the following matrix: take the identity matrix of size n, then replace the i-th row by the i-th row of the $A(G)$, divided by the sum of the elements in the row. More formally, we have that for any row $p_{j*} \in P_i$,

$$p_{j*} = \begin{cases} I_{j*} & j \neq i \\ \frac{1}{\sum_{k=1}^{n} A(G)_{jk}} A(G)_{j*} & j = i \end{cases}$$

It is quite easy to observe that any P_i has rows that sum to 1, and therefore it is right stochastic. Additionally, applying P_i to a vector representing the current state of values in G will produce a vector that corresponds to the values in G after node i takes on the average of its neighbors' values.

Let $V = \prod_{t=1}^{T} P_{i(t)}$, where $i(t)$ is the index of the node that is undergoing the averaging process at time t. The problem can thus be rephrased as

$$V\vec{g} = \mu \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

where μ is the final consensus value that the process converges on, and is dependent on both the input vector $\vec{g}$ as well as the ordering of the P in V. Therefore, when we change V, we also change μ in most cases, and when we say "node $x = \mu$" what is really meant is that node x has reached the consensus state implied by V at that particular point in time (when perhaps the entire process has not finished yet). What we want to find is a condition on G such that minimum T required for the process to terminate is finite.

Note that in this formulation, nodes only take their averages one by one, which means that two adjacent nodes cannot undergo the process at the same time, unlike in the original problem.

3 Finite Time Consensus

Proposition 1. *$\exists\, V$ such that T is finite for all $\vec{g}$ iff G is a complete bipartite graph.*

We begin with the backward implication, which shows that there exists a sequence of averaging processes which makes all the values in a complete bipartite graph the same. It is quite obvious that by splitting the nodes into the two disjoint independent sets α and β, and by first taking the process over members of one part then the other, we arrive at a consensus. In this case, whichever set came afterwards, be it α or β, will have its average value be the consensus value.

At this point it should be noted that for any sequence of P, we can write $P_\alpha = \prod_{a\in\alpha} P_a$ if α is an independent subset of the vertices. Therefore, we can express V in the bipartite graph as $V = P_\alpha P_\beta$ or $P_\beta P_\alpha$. Observe that the product of P over vertices in an independent set does not depend on the order in which the averages are taken over the vertices, since the value of each does not affect any other node's value from the set.

Now we consider the forward implication, where I attempt to prove that complete bipartite graphs are the only structures on which this process achieves consensus in a finite number of steps. This is an attempt that relies on a heuristic argument at one point, which I have not found a satisfactory way of proving rigorously.

First, consider the following few graphs.

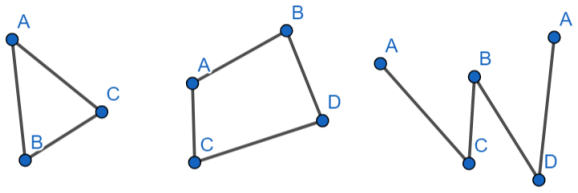

Figure 2: Examples of graphs with variable starting values. The left graph will not reach consensus in finite time, but the middle and right graphs will. Changing one of the nodes with value A in the right graph will also make it unable to reach consensus in finite time.

In the left graph, there is no finite series of P that allows the values to converge every time. In the middle graph, we can simply take $\alpha = \{A, D\}$ and $\beta = \{B, C\}$, and it will reach a consensus state as it is a complete bipartite graph. The graph on the right is bipartite but not complete. If one of the nodes labeled A had a different value, such as E, then the system will not reach

consensus. Here, because both of the end nodes are labeled A, we can consider this graph to be "equivalent" to the middle graph where one of its nodes is split into two.

In general, if a graph has $|G|$ vertices and d of them need to be deliberately chosen for consensus to be reached in finite time, then the graph has $|G| - d$ degrees of freedom. The graphs above have degrees of freedom equaling 2, 4, and 4 respectively.

The heuristic argument invoked here is that for any graph which permits a finite time consensus, there exists V such that $T \leq |G| - d$. Another way of thinking about this is that at every point in time during the process, it is turning the value in a specific node into μ, and this node has never been visited before, nor will it be visited ever again by the process. In other words, there are no repeated P_i in the expansion of V, and hence there is no scenario in which the values of two nodes i and j at different times in the process, which we denote as $v_i^t, v_j^{t'}$, and $v_i^{t''}, t > t' > t''$, are related in a way such that $v_i^t = \mu$ depends on $v_j^{t'} = \mu$ which depends on $v_i^{t''} = \mu' \neq \mu$.

Unfortunately, this is only a heuristic argument based on observations of structures. If there is a repeated P_i in the process, then between the two copies of P_i there must be a P_j where j and i are adjacent. This is because P for non-adjacent cells commute and every P_i is idempotent, so without P_j one of the two copies will just cancel out. Then since the earlier P_i was considered necessary, it depended on $v_j^{t''}$, which means that if $v_j^{t''} \neq v_i^{t''}$ but $v_j^t = v_i^t$, then the averaging process discretely converged to a stable point despite being a continuous process where we expect decreasing incremental moves towards the consensus value. Therefore, unless the changes to v_i and v_j after time t'' were unnecessary, it is reasonable to believe that the process here is not finite, as it is converging towards μ in a limit-like fashion.

The proof of the first question raised in the introduction only relies on the validity of the above heuristic in the case where $d = 0$, whereas the second question relies on the generalization. It is possible that a rigorous proof for this argument requires an analysis of the finite time stabilization of Lyapunov functions defined over the set of graph values, an example of which would be the variance of the values, but that is beyond my ability. Thus, I take it for granted that this assumption is true, since I have no reason to believe otherwise.

Now, assume that G is a graph with $|G|$ degrees of freedom, for which we have defined $V = P_D P_\alpha P_\beta$, where β is an independent set and α is an independent set whose members are all adjacent to exactly the same members of β. This form is guaranteed to exist in every V since we can just take the rightmost

(i.e. earliest) P up until the first node which is adjacent to at least one of the previously processed nodes, let that set be β, and let the first adjacent node be the only member of α (though more can be added). A bit of manipulation shows that the principal submatrix where the indices correspond to the elements of $\alpha \cup \beta$ has rank 1 (though rows are not guaranteed to be exactly the same, even though $P_\alpha P_\beta$ is still right stochastic overall). Furthermore, in V, the columns corresponding to elements of β consist only of zeroes.

We can reindex the nodes for convenience such that the elements of $\alpha \cup \beta$ are on the bottom right of V. Then, we can express $P_\alpha P_\beta$ as

$$\begin{bmatrix} I_{|G|-|\alpha\cup\beta|} & 0 \\ R & V' \end{bmatrix}$$

where V' is the rank 1 submatrix. Now express P_D similarly as

$$\begin{bmatrix} H & M \\ N & W \end{bmatrix}$$

such that we have

$$V = \begin{bmatrix} H + MR & MV' \\ N + WR & WV' \end{bmatrix}$$

The heuristic assumption from above allows us to assume that $W = I_{|\alpha\cup\beta|}$, and as a result we also have $N = 0$. Since we know that V is right stochastic and has rank 1, all of its rows must be the exact same. This then suggests that all the rows in $[R\ V']$ are the exact same.

Now if we focus on MV', we see that all of its rows must be the same as the rows in V'. Thus, for any row m_{i*} in M, we have

$$
\begin{aligned}
m_{i*}V' &= m_{i*} \cdot \begin{bmatrix} v'_{*1} & v'_{*2} \dots v'_{*(|\alpha \cup \beta|)} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{j=1}^{|\alpha\cup\beta|} m_{ij}v'_{j1} & \sum_{j=1}^{|\alpha\cup\beta|} m_{ij}v'_{j2} \dots & \sum_{j=1}^{|\alpha\cup\beta|} m_{ij}v'_{j(|\alpha\cup\beta|)} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{j=1}^{|\alpha\cup\beta|} m_{ij}k_1 & \sum_{j=1}^{|\alpha\cup\beta|} m_{ij}k_2 \dots & \sum_{j=1}^{|\alpha\cup\beta|} m_{ij}k_{|\alpha\cup\beta|} \end{bmatrix} \\
&= \begin{bmatrix} k_1 & k_2 \dots k_{|\alpha\cup\beta|} \end{bmatrix} \text{ by definition.} \\
\therefore \sum_{j=1}^{|\alpha\cup\beta|} m_{ij} &= 1
\end{aligned}
$$

Therefore, since all values in V are non-negative, it must be the case that $H = 0$. Thus, we now have that

$$P_D = \begin{bmatrix} 0 & M \\ 0 & I \end{bmatrix}$$

From our prior analysis, we know that this represents the effect of the process on an independent set, and it just so happens that this independent set $D = G\backslash(\alpha \cup \beta)$. However, since the values of nodes in α and β have reached μ before P_D is applied and do not change, this means that all the nodes in $\alpha \cup \beta$ must be adjacent to the exact same set of nodes in D, else changing some subset of initial values in D will cause α and β to no longer reach consensus in the beginning. We also know that $\forall\, d \in D$, it is adjacent to at least one $a \in \alpha$, else it should have been "absorbed" into α in the first place. Combining these two propositions results in the observation that D is an independent set whose elements are connected to every single node in $\alpha \cup \beta$. Additionally, since β is fixed before the application of P_α, a similar line of logic suggests that all elements of α are connected to every element of β.

At this point, we have three independent sets, and in every set each element is adjacent to every element of the sets which it is not a part of. Let the averages of the initial values of each set be μ_α, μ_β, and μ_D. After P_β is applied, we can already calculate the consensus value:

$$\mu = \frac{|\alpha|\mu_\alpha + |D|\mu_D}{|\alpha| + |D|}$$

And after P_α:

$$\mu = \frac{|D|\mu_D + |\beta|\frac{|\alpha|\mu_\alpha + |D|\mu_D}{|\alpha|+|D|}}{|\beta| + |D|}$$

Combining the two we get

$$\begin{aligned}
\frac{|\alpha|\mu_\alpha + |D|\mu_D}{|\alpha| + |D|} &= \frac{|D|\mu_D + |\beta|\frac{|\alpha|\mu_\alpha + |D|\mu_D}{|\alpha|+|D|}}{|\beta| + |D|} \\
|\alpha|\mu_\alpha + |D|\mu_D &= \frac{(|\alpha| + |D|)|D|\mu_D + |\beta|(|\alpha|\mu_\alpha + |D|\mu_D)}{|\beta| + |D|} \\
|D|(|\alpha|\mu_\alpha + |D|\mu_D) &= (|\alpha| + |D|)|D|\mu_D \\
\mu_\alpha &= \mu_D
\end{aligned}$$

This contradicts the fact that G has $|G|$ degrees of freedom, and thus we cannot define V such that our process both ends in finite time T and admits any random initial input $\vec{g}$.

4 Second Problem, Generalizations, and Non-Heuristic Proofs

In the original question, it was assumed that adjacent cells can simultaneously update their values. In fact, this is the same as the current formulation where adjacent cells cannot run the process simultaneously if we take the heuristic argument to be true. Since the convergence towards μ for any node is dependent on some of its previous nodes already reaching μ (or if it is in β), then running the process on adjacent nodes simultaneously will not change whether a node reaches consensus as compared to the sequential version. Inversely, it is obvious that a sequential algorithm can run in the simultaneous version. Therefore, the two problems are the same insofar as the heuristic argument is assumed to be true.

The difficult part comes down to proving the heuristic argument in full. There are some possibilities for its proof that could be explored. First would be the consideration of v_i and v_k, where i and k have lower and greater values respectively compared to j. Some iteration of this logic on the neighbors of k reveals that some k' must exist where all its neighbors depend on $k \neq \mu$ to equal μ. From here, it seems likely that there must be some contradictory outcome when the neighbors of the neighbors of k' are considered, unless the graph is

not finite. Another possibility is to investigate the properties of the finite set from which nodes draw their values during the process. It seems unlikely that any conclusion about this problem can be derived from further investigation of P_i.

For the second problem, there is an upper bound on the d given by the result to the first problem. We only need to find the largest independent set $E \subset G$, and fix all the other nodes to μ, so that $d \leq |G \backslash E|$. Whether this is the optimal value remains to be seen, though it is highly likely that it is a tight bound.