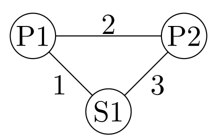
EN.580.694 ASSIGNMENT # 06

HEATHER G. PATSOLIC
STATISTICAL CONNECTOMICS

In Fino 2011, they pose a question regarding connection selectivity between sGFP and PC cells (can find the paper in the google drive folder if you forget). What your homework for Tuesday 4/13 is, is the following:

- type up a 1 page document suggesting how you would redefine our graph from Fino presented in class (i.e. redefine nodes and edges) OR
- type up a 1 page document arguing why you think that in this case it is not possible using graph theory approaches.

1. The Question At Hand



Th question we would like to answer is whether or not $P(3|(2 \cap 1)) = P(3|1)$. We want to be able to do some sort of hypothesis testing where the null hypothesis is that these two quantities are equivalent and the alternative hypothesis is that they are not. However, in order to determine this, we need a model for the hypothesis. We want to create a graphical representation of this problem. Our sample space is still $\Omega = \{n, E, Y\}$ where n is the set of vertices, with E is the set of edges and Y is a vector which holds a label for each vertex as to whether or not it represents an sGFP or a PC cell.

2. Conditional Interference

I am not sure that there is a good model for this question. At least not a simple one. The main challenge I have been finding is in constructing a model which lends itself to finding two different conditional probabilities. We want to determine if the probability of edge 3 given edges 2 and 1 is the same as the probability of edge 3 given just edge 1. So, we need to construct a graphical model in which we can easily compare the proportion of triangles to $P1-P2-S1$ and $P2-P1-S1$ lines and also be able to easily compare the proportion of $P1-S1$ and $P2-S1$ edges to $P1-S1-P2$ lines.

3. A Few Misguided Ideas

Below, I describe a few ideas I had about how to model this problem graphically that ended up having problems that I have not figured out how to address.

(1) At first, it seemed that maybe we want to determine the proportion of triangles (as shown above) as opposed to lines $P1-S1-P2$. In order to do this, I originally thought of using the dual graph where there is a vertex representing each possible triangle $P1-S1-P2-P1$ with edges connecting these vertices if and only if they share a common edge. In this way, we would find that the degree 3 vertices represented the triangles we are interested in. The trouble comes in identifying the degree 2 vertices. A vertex in this graph could have degree

Date: April 14, 2015.

two because of sharing edges $P1 - S1$ and $S1 - P2$ or because of sharing edges $S1 - P1$ and $P1 - P2$ or $S1 - P2$ and $P2 - P1$.

(2) Another thought was to compare two different models and see if they have the same proportions. That is, can we create one graphical model to represent finding an edge 3 where edges 2 and 1 are given in the construction and another graph model to represent finding an edge 3 where edge 1 is given in the construction. We would then determine the proportion of edge 3 in the first and the proportion of edge 3 in the second and see if these two quantities were essentially equivalent via hypothesis testing.

For the first, we could create a bipartite graph where the PC cells are collapsed if they share an edge. That is, for each pair of PC cells, determine if they are adjacent, if so then create a vertex representing this pair of cells, if not then do not create a vertex. We would also have a vertex representing each sGFP cell that is adjacent to at least one PC cell and create edges between an sGFP cell and a PC-pair if and only if the sGFP was adjacent to both PC cells in the pair. With this model, we could determine the number sGFP cells connected to pairs of PC cells and compare to the number of sGFP cells present in the graph; however, there would be error in this proportion as an estimate for $P(3|1 \cap 2)$ in that we would only be counting each sGFP cell once and it could be attached to more than one PC cell across different pairs of PC cells. For example if we have pairs $P(1/2)$ and $P(2/3)$ then sGFP could be connected to $P1$ and $P3$ and we would get an over-estimate of the proportion by counting the number of $sGFP$ cells as opposed to the number of $P1 - S1$ connections.

For the second, perhaps we could create a graph where a vertex is created for every edge which is adjacent to exactly one sGFP cell and one PC cell. That would give the total number of these connections. We could then create an edge between two cells if and only if they have the same sGFP cell and different PC cells. Since in the second graph we do not care about $P1 - P2$ adjacencies, we ignore them. We could then use the number of edges divided by the number of vertices as an estimate for $P(1|3)$.