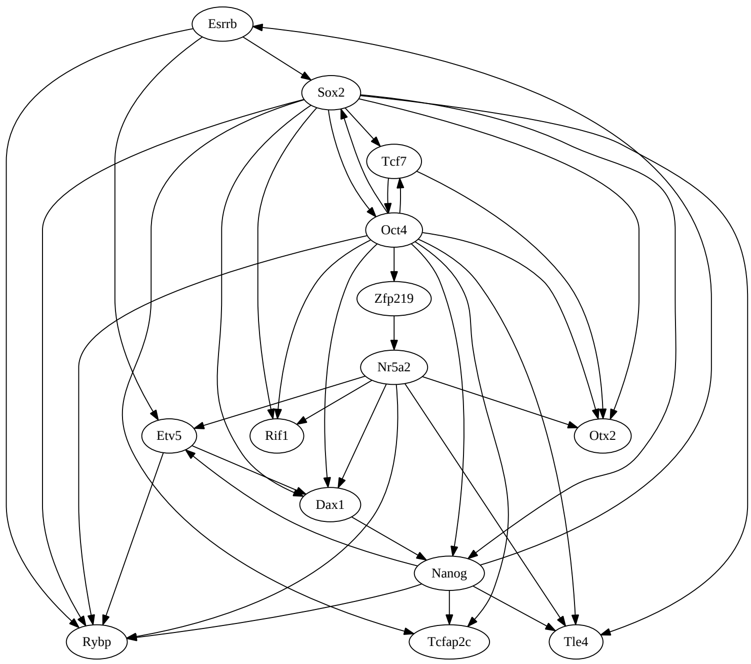
# MS6a, Exercises Week 3

Rune Lyngsø

October 28, 2009

## A Strongly Connected Components

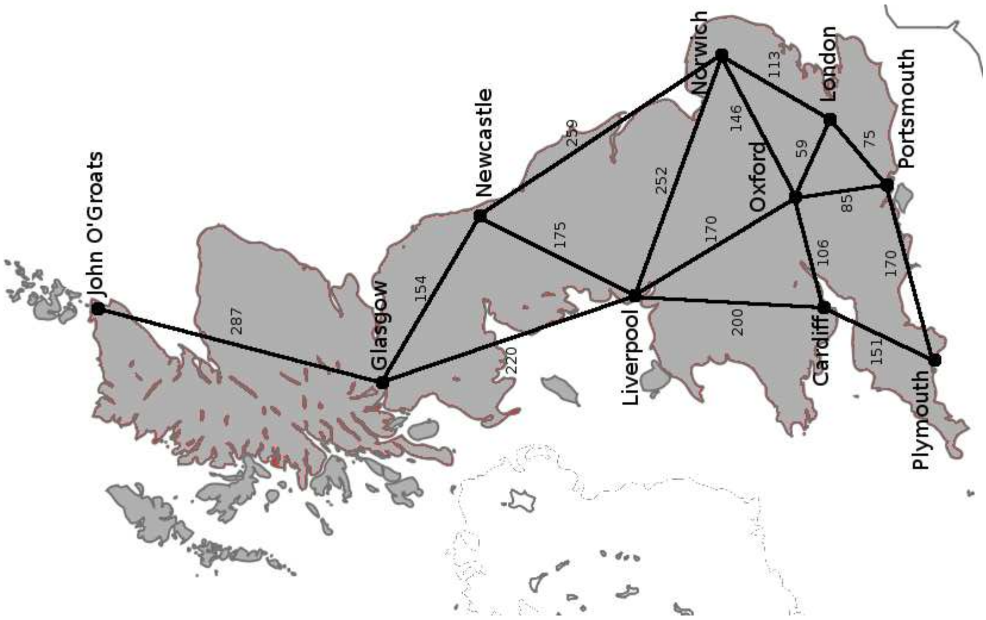
1. Find the strongly connected components in the following regulatory network, by running the strongly connected component algorithm in the lecture notes. For each gene, include the finishing time for the corresponding node in your answer, and indicate which gene leads to the discovery of each strongly connected component. This network is a subnetwork of the network presented in [1, Fig. 3], with two regulatory interactions changed to have opposite direction (Zfp219 → Nr5a2 and Dax1 → Nanog) and one interaction removed (Nanog → Oct4).




2. The equivalent to strongly connected components for an undirected graph can be said to be *biconnected components*. Two edges are biconnected if they lie on a common simple cycle (i.e. a path starting and ending in the same node, but otherwise consisting of distinct nodes). A biconnected component cannot become disconnected by deletion of a single node or edge. Note that a node can be incident to edges in different biconnected components. Describe an algorithm that finds all biconnected components by a liberal use of maximum flow problem solutions.

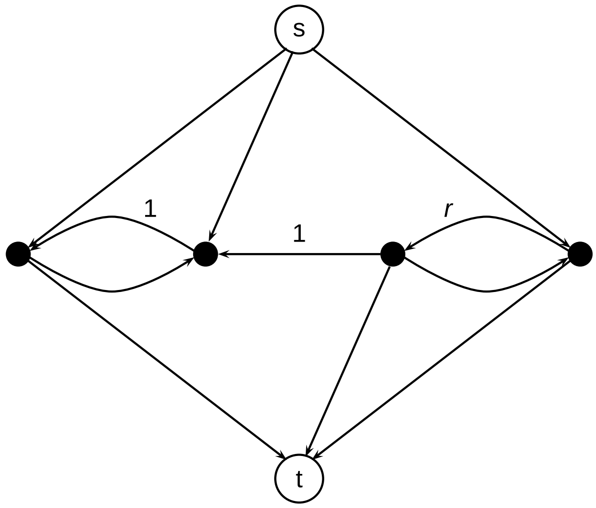
## B Maximum Flows

3. In the road network from problem 1 on last week's set of exercises,



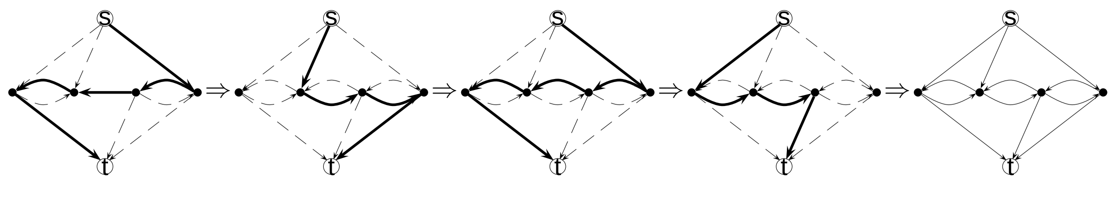

assume that edge weights are capacities, and that there are edges with this capacity in both directions of the edge. What is the maximum flow from Newcastle to Plymouth? What is the corresponding minimum capacity cut?

4. Consider the flow network




with all edges, except for the three annotated with capacities 1, 1, and $r$, having capacity $M$ for some suitably large value of $M$. Further assume that $r$ has been chosen such that $r = \frac{1}{1+r}$, i.e. $r$ is the golden ratio.

Assume that we make four flow updates with the maximum capacity of the following augmenting paths (paths shown in bold, only horizontal back edges in the residual network shown):



What are the residual capacities of the three edges with original capacities 1, 1, and $r$? Show that $1 - r = r^2$ and $2r - 1 = r^3$. Argue that repeating this chain of augmenting path updates will never generate a maximum flow. What is the maximum flow of the network?

5. Assume $G = (V, E)$ describes a metabolic network, with $V$ being the set of compounds and directed edges $E$ being the set of reactions converting a reactant into a product. How can you determine the minimum number of reactions that need to be disabled before we can no longer produce a target product $t \in V$ from initial metabolite $s \in V$?

6. Assume $G = (V, E)$ describes a regulatory network, with $V$ being the set of genes and directed edges $E$ indicating that one gene regulates another. How can you determine the minimum number of genes that need to be deleted for a gene $s \in V$ not to have any regulatory effect on another gene $t \in V$?

7. Consider the alternative splicing graph of problem 4 of last week's exercise,

| | | | | |
|---|---|---|---|---|
| $0 \to \{1, 3, 24\}$ | $1 \to \{2, 4\}$ | $2 \to \{5\}$ | $3 \to \{4\}$ | $4 \to \{5\}$ |
| $5 \to \{6, 7, 8, 11\}$ | $6 \to \{7\}$ | $7 \to \{8\}$ | $8 \to \{9, 11\}$ | $9 \to \{10, 11\}$ |
| $10 \to \{11\}$ | $11 \to \{12, 13, 16\}$ | $12 \to \{13, 14, 15\}$ | $13 \to \{14\}$ | $14 \to \{15\}$ |
| $15 \to \{16\}$ | $16 \to \{17, 18, 20\}$ | $17 \to \{18\}$ | $18 \to \{19, 20, 21, 22, 39\}$ | $19 \to \{20\}$ |
| $20 \to \{21\}$ | $21 \to \{22, 23, 29\}$ | $22 \to \{23\}$ | $23 \to \{25, 26, 32\}$ | $24 \to \{25\}$ |
| $25 \to \{26, 28\}$ | $26 \to \{27, 31, 38\}$ | $27 \to \{28\}$ | $28 \to \{29\}$ | $29 \to \{30, 31\}$ |
| $30 \to \{31\}$ | $31 \to \{32\}$ | $32 \to \{33, 34, 35\}$ | $33 \to \{35\}$ | $34 \to \{35\}$ |
| $35 \to \{36, 37, 39\}$ | $36 \to \{37\}$ | $37 \to \{38\}$ | $38 \to \{39\}$ | $39 \to \{40\}$ |
| $40 \to \{41, 42\}$ | $41 \to \{42\}$ | $42 \to \{43\}$ | $43 \to \emptyset$ | |

where for each node the set of nodes it has an edge to is listed above. As discussed, a path through this graph corresponds to a spliceform that can be observed, i.e. a possible transcript we may sequence. Hence, the minimum number of transcripts we need to sequence before we can have any hope of having seen all edges in the alternative splice graph is equal to the minimum number of paths required to make sure that every edge is in at least one path.

An alternative splicing graph is always acyclic. Given such a graph $G = (V, E)$, define a flow network $G' = (V', E', w')$ with

$$V' = \{x_u\}_{u \in V} \cup \{y_u\}_{u \in V} \cup \{s, t\}$$
$$E' = \{(s, x_u)\}_{u \in V} \cup \{(y_u, t)\}_{u \in V} \cup \{(x_u, y_v) \mid (u, v) \in E\}$$

and all edge capacities set to 1. What is the relationship between the minimum number of paths needed to cover $G$, and the maximum flow in $G'$? What is the minimum number of transcripts that we need to sequence to detect all edges in the alternative splicing graph above?

Assume that we have an alternative splicing model where for each node we have a probability over the outgoing edges. Can you describe a recursion allowing us to compute the probability that all edges are observed if we draw $k$ transcripts from the splicing model? (hint: recurse on nodes according to the natural ordering, and keep track of the number of partial paths currently terminating at each node up to the current node).

## References


[1] Q. Zhou, H. Chipperfield, D. A. Melton, and W. H. Wong. A gene regulatory network in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 104(42):16438–16443, 2007.