BIOINFORMATICS
Vol. 00 no. 00 2010
Pages 1–5

A Poisson Model for Random Multigraphs

John M. O. Ranola [1], Sangtae Ahn [2], Mary Sehl [1], Desmond J. Smith [3], and Kenneth Lange [1,4,5]*

[1]Department of Biomathematics, University of California Los Angeles
[2]Department of Electrical Engineering, University of Southern California
[3]Department of Molecular and Medical Pharmacology, University of California Los Angeles
[4]Department of Human Genetics, University of California Los Angeles
[5]Department of Statistics, University of California Los Angeles

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

APPENDIX

Existence and Uniqueness of the Estimates

The body of the paper takes for granted the existence and uniqueness of the maximum likelihood estimates. These more subtle questions can be tackled by reparameterizing. Before we do so, let us dismiss the exceptional cases where a node has no edges. If this condition holds for node i, then in the undirected graph model the value $p_i = 0$ maximizes $L(\mathbf{p})$ regardless of the values of the other parameters p_j. In the directed graph model, if node i has no outgoing arcs, then likewise we should take $p_i = 0$, and if i has no incoming arcs, then we should take $q_i = 0$.

The reparameterization we have in mind is $p_i = e^{r_i}$ and $q_i = e^{s_i}$. It is clear that the reparameterized loglikelihoods

$$L(\mathbf{r}) = \sum_{\{i,j\}} [x_{ij}(r_i + r_j) - e^{r_i+r_j} - \ln x_{ij}!] \quad (1)$$

$$L(\mathbf{r}, \mathbf{s}) = \sum_i \sum_{j \neq i} [x_{ij}(r_i + s_j) - e^{r_i+s_j} - \ln x_{ij}!] \quad (2)$$

are concave. If an original parameter p_i is set to 0, then we drop all terms from the loglikelihood involving r_i. If there are only two nodes, then the loglikelihood $L(\mathbf{r})$ is constant along the line $r_1 + r_2 = 0$. In the directed graph model, if an original parameter q_j is set to 0, then we drop all terms from the loglikelihood involving s_j. With only two nodes, the loglikelihood $L(\mathbf{r}, \mathbf{s})$ is constant on the subspace defined by the equations $r_1 + s_2 = 0$ and $r_2 + s_1 = 0$. Strict concavity and uniqueness of the maximum likelihood estimates fail in each instance. Thus, assume that the number of nodes $m \geq 3$.

For strict concavity to hold, the positive semidefinite quadratic form

$$-v^t d^2 L(\mathbf{r})\mathbf{v} = \sum_{\{i,j\}} (v_i + v_j)^2 e^{r_i+r_j}$$

must be positive definite. When the quadratic form vanishes, $v_i + v_j = 0$ for all pairs $\{i, j\}$. If some $v_i \neq 0$, then $v_j = -v_i \neq 0$ for all $j \neq i$. With a third node k distinct from i and j, we have $v_j + v_k = -2v_i \neq 0$. This contradiction shows that $\mathbf{v} = \mathbf{0}$ and proves that $L(\mathbf{r})$ is strictly concave. It follows that there can be at most one maximum point.

In the directed graph model, it is clear that we can replace each r_i by $r_i + c$ and each s_j by $s_j - c$ without changing the value of the loglikelihood (1). In other words, the loglikelihood is flat along a line segment, and strict concavity fails. If we impose the constraint $r_1 = 0$ corresponding to $p_1 = 1$, then things improve. Consider the semipositive definite quadratic form

$$-\mathbf{w}^t d^2 L(\mathbf{r}, \mathbf{s})\mathbf{w} = \sum_i \sum_{j \neq i} (u_i + v_j)^2 e^{r_i+s_j}$$

where $\mathbf{w}$ equals the concatenation of the vectors $\mathbf{u}$ and $\mathbf{v}$. The constraint $r_1 = 0$ corresponding to $p_1 = 1$ allows us to drop the variable u_1, and the term $(u_1 + v_j)^2 e^{r_1+s_j}$ of the quadratic form becomes $v_j^2 e^{s_j}$. In order for the quadratic form to vanish, we must have $v_j = 0$ for all j. This in turn implies that all u_i must vanish for $i \neq 1$. Hence, $L(\mathbf{r}, \mathbf{s})$ is strictly concave under the proviso that $r_1 = 0$, and again we are entitled to conclude that at most one maximum point exists.

Existence rather than uniqueness of a maximum point depends on the property of coerciveness summarized by the requirement $\lim_{\|\mathbf{r}\| \to \infty} f(\mathbf{r}) = \infty$ for the convex function $f(\mathbf{r}) = -L(\mathbf{r})$. Equivalently, each of the sublevel sets $\{\mathbf{r} : f(\mathbf{r}) \leq c\}$ is compact. For a convex function $f(\mathbf{r})$, coerciveness is determined by the asymptotic function

$$f'_\infty(\mathbf{d}) = \sup_{t>0} \frac{f(t\mathbf{d}) - f(\mathbf{0})}{t} = \lim_{t \to \infty} \frac{f(t\mathbf{d}) - f(\mathbf{0})}{t}.$$

A necessary and sufficient condition for all sublevel sets of $f(\mathbf{r})$ to be compact is that $f'_\infty(\mathbf{d}) > 0$ for all vectors $\mathbf{d} \neq \mathbf{0}$ (Hiriart-Urruty, 2004). In the present circumstances,

$$f'_\infty(\mathbf{d}) = \sup_{t>0} \sum_{\{i,j\}} \Big[\frac{e^{t(d_i+d_j)} - 1}{t} - x_{ij}(d_i + d_j)\Big].$$

If any sum $d_i + d_j > 0$, then it is obvious that $f'_\infty(\mathbf{d}) > 0$. Thus, we may assume that all pairs satisfy $d_i + d_j \leq 0$. With this assumption in place, if some $x_{ij} > 0$, then the assumption $d_i +$

*to whom correspondence should be addressed

$d_j < 0$ also gives $f'_\infty(\mathbf{d}) > 0$. Hence, we may also assume that $d_i + d_j = 0$ for all pairs with $x_{ij} > 0$. If all $d_j \leq 0$, suppose $d_i < 0$. Then there is at least one j with $x_{ij} > 0$. But this entails $d_i + d_j = 0$ and contradicts our assumption that $d_j \leq 0$. Finally, let us assume some $d_i > 0$. Then $d_j < 0$ for all $j \neq i$. If $x_{jk} > 0$ for a pair $\{j, k\}$ with $j \neq i$ and $k \neq i$, then $d_j + d_k = 0$ and either d_j or d_k is positive. This is a contradiction. Hence, all edges involve i. Because all nodes lacking edges are omitted from consideration, all nodes are connected to i. In other words, the only way the condition $f'_\infty(\mathbf{d}) > 0$ can occur with $\mathbf{d} \neq \mathbf{0}$ is for i to serve as a hub in the narrow sense of attracting all edges.

A hub formation is incompatible with coerciveness. Indeed, suppose i is the hub. If we take $r_i = t > 0$ and all $r_j = -t$ for $j \neq i$, then the loglikelihood (1) becomes

$$L(\mathbf{r}) = \sum_{j \neq i} [x_{ij}(t - t) - e^{t-t} - \ln x_{ij}!] - \sum_{\{j,k\}:j \neq i, k \neq i} e^{-2t},$$

which is bounded below as $t \to \infty$. A two-node model obviously involves two hubs.

Hubs also supply the only exceptions to coerciveness in the directed graph model. In proving this assertion, we let I be the set of nodes with incoming arcs and O be the set of nodes with outgoing arcs. The parameter r_i is defined provided $i \in O$, and the parameter s_j is defined provided $j \in I$. Suppose i is a hub with both outgoing and incoming arcs. Set $r_i = 0$, $s_i = t$, $s_j = 0$ when $j \in I \setminus \{i\}$, and $r_j = -t$ when $j \in O \setminus \{i\}$. The loglikelihood

$$\begin{aligned} L(\mathbf{r}, \mathbf{s}) &= \sum_{j \in I \setminus \{i\}} [x_{ij} 0 - e^0 - \ln x_{ij}!] \\ &+ \sum_{j \in O \setminus \{i\}} [x_{ji}(-t + t) - e^{-t+t} - \ln x_{ij}!] \\ &- \sum_{j \in O \setminus \{i\}} \sum_{k \in I \setminus \{i,j\}} e^{-t} \end{aligned}$$

remains bounded as t tends to ∞. Thus, $L(\mathbf{r}, \mathbf{s})$ fails to be coercive in this setting.

In proving the converse for a directed graph, we write the asymptotic function as

$$f'_\infty(\mathbf{c}, \mathbf{d}) = \sup_{t>0} \sum_{i \in O} \sum_{j \in I \setminus \{i\}} \Big[\frac{e^{t(c_i + d_j)} - 1}{t} - x_{ij}(c_i + d_j) \Big].$$

A pair (i, j) is said to be active provided $i \in O$ and $j \in I$. If the loglikelihood is not coercive, then there exists a vector $(\mathbf{c}, \mathbf{d}) \neq \mathbf{0}$ with $f'_\infty(\mathbf{c}, \mathbf{d}) = 0$, where $\mathbf{c}$ is the vector of defined c_i and d is the vector of defined d_j. It suffices to show that $f'_\infty(\mathbf{c}, \mathbf{d}) = 0$ for some nontrivial $(\mathbf{c}, \mathbf{d})$ is impossible unless the graph is organized as a hub with both incoming and outgoing arcs.

Without loss of generality, we can assume that $x_{12} > 0$; otherwise, we relabel the nodes so that some arc starts at node 1 and ends at node 2. This choice also allows us to eliminate the propensity r_1 and set $c_1 = 0$. If $c_i + d_j > 0$ for an active pair (i, j), then it is obvious that $f'_\infty(\mathbf{c}, \mathbf{d}) > 0$. Furthermore, if $x_{ij} > 0$ and $c_i + d_j < 0$, then we also have $f'_\infty(\mathbf{d}) > 0$. Thus, we may assume that all active pairs (i, j) satisfy $c_i + d_j \leq 0$, with equality when $x_{ij} > 0$. Given these restrictions, the assumption $c_1 = 0$ requires that $d_j \leq 0$ for all $j \neq 1$ in I. In view of our assumption $x_{12} > 0$, we find that $d_2 = 0$. If $k \neq 2$ is in O, the restriction $c_k + d_2 \leq 0$ implies that $c_k \leq 0$. Thus, the only two components that can be positive are d_1 and c_2. Suppose the pair $(2, 1)$ is active. The inequality $c_2 + d_1 \leq 0$ implies that if either component d_1 or c_2 is positive, then the other component is negative. Similarly, if $x_{kl} > 0$ for nodes $k \neq 2$ and $l \neq 1$, then the equality $c_k + d_l = 0$ and the nonpositivity of c_k and d_l yield $c_k = d_l = 0$.

If we can show that c_2 and d_1 are nonpositive when defined, then all components of $(\mathbf{c}, \mathbf{d})$ will be nonpositive. This state of affairs actually implies that all components are 0, contradicting our assumption that $(\mathbf{c}, \mathbf{d})$ is nontrivial. To prove this claim, consider a defined component c_i. Because there exists a node j with $x_{ij} > 0$, the equation $c_i + d_j = 0$ entails $c_i = 0$ when all components of $(\mathbf{c}, \mathbf{d})$ are nonpositive. Likewise, for every defined d_j, there exists a node i with $x_{ij} > 0$. The equation $c_i + d_j = 0$ now entails $d_j = 0$ when all components of $(\mathbf{c}, \mathbf{d})$ are nonpositive.

The proof now separates into cases. In the first case, no other arcs impinge on node 1 or node 2 except possibly the arc $2 \to 1$. If the arc $2 \to 1$ does not exist, d_1 and c_2 are undefined, and we are done. If $2 \to 1$ exists, then to a avoid a hub with both incoming and outgoing arcs, there must be a third arc $k \to l$ distinct from $1 \to 2$ and $2 \to 1$. We have already observed that $c_k = d_l = 0$ for an arc $k \to l$ with $k \neq 2$ and $l \neq 1$. Therefore, the requirement $c_k + d_1 \leq 0$ entails $d_1 \leq 0$. Similarly, the requirement $c_2 + d_l \leq 0$ entails $c_2 \leq 0$.

In the second case, component d_1 is defined and component c_2 is undefined. To prevent node 1 from being a hub with both incoming and outgoing arcs, there must be an arc $k \to l$ with k and l different from 1. Because c_2 is undefined, $k \neq 2$. Hence, again $c_k = d_l = 0$. The requirement $c_k + d_1 \leq 0$ now implies $d_1 \leq 0$.

In the third case, component d_1 is undefined and component c_2 is defined. To prevent node 2 from being a hub with both incoming and outgoing arcs, there must be an arc $k \to l$ with k and l different from 2. Because d_1 is undefined, $l \neq 1$. Hence, again $c_k = d_l = 0$. The requirement $c_2 + d_l \leq 0$ now implies $c_2 \leq 0$.

In the fourth and final case, both components d_1 and c_2 are defined. The hub hypothesis fails if there exists an arc $k \to l$ with k and l both differing from 1 and 2. As noted earlier, this leads to the conclusions $d_1 \leq 0$ and $c_2 \leq 0$. If no such arc exists, then consider arcs $k \to 1$ and $2 \to l$. If the only possible k is $k = 2$, then node 2 is a hub with both incoming and outgoing arcs. Assuming $k \neq 2$, we have $c_k \leq 0$. The requirement $c_k + d_1 = 0$ now implies $d_1 \geq 0$. In similar fashion, if the only possible value of l is 1, then node 1 is a hub with both incoming and outgoing arcs. Assuming $l \neq 1$, we have $d_l \leq 0$. The requirement $c_2 + d_l = 0$ now implies $c_2 \geq 0$. Unless $d_1 = c_2 = 0$, the two conditions $d_1 \geq 0$ and $c_2 \geq 0$ are incompatible with our earlier finding that $d_1 > 0$ implies $c_2 < 0$ and vice versa.

In summary, we have found that the condition $f'_\infty(\mathbf{c}, \mathbf{d}) = 0$ and the assumption of no hub with both incoming and outgoing arcs imply that $(\mathbf{c}, \mathbf{d}) = \mathbf{0}$. Thus, the strictly convex function $f(\mathbf{r}, \mathbf{s}) = -L(\mathbf{r}, \mathbf{s})$ is coercive under the no hub assumption and attains its maximum at a unique point.

Convergence of the MM Algorithms

Verification of global convergence of the MM algorithms hinge on five properties of the objective function $L(\mathbf{p})$ and the iteration map $M(\mathbf{p})$:

(a) $L(\mathbf{p})$ is coercive,

(b) $L(\mathbf{p})$ has only isolated stationary points,

(c) $M(\mathbf{p})$ is continuous,

(d) A point is a fixed point of $M(\mathbf{p})$ if and only if it is a stationary point of $L(\mathbf{p})$,

(e) $L[M(\mathbf{p})] \leq L(\mathbf{p})$, with equality if and only if $\mathbf{p}$ is a fixed point of $M(\mathbf{p})$.

See the reference (Lange, 2004) for full details.

Verification of these properties in the multigraph models is straightforward. Coerciveness has already been dealt with under the reparameterization $p_i = e^{r_i}$ and the no hub assumption. Because the reparameterized loglikelihood $L(\mathbf{r})$ is strictly concave, there is a single stationary point in both the original and transformed coordinates. Inspection of the iteration map (Equation 3 in paper) shows that it is continuous. It does involve a division by a denominator that could tend to 0, but this contingency is ruled out by coerciveness. The fixed point condition $M(\mathbf{p}) = \mathbf{p}$ occurs when the surrogate function satisfies the equation $\nabla g(\mathbf{p} \mid \mathbf{p}) = \mathbf{0}$. The identity $\nabla L(\mathbf{p}) = \nabla g(\mathbf{p} \mid \mathbf{p})$ at every interior point of the domain of the objective function shows that fixed points and stationary points coincide. Finally, the strict concavity of the surrogate function $g(\mathbf{p} \mid \mathbf{p}^n)$ demonstrates that $g(\mathbf{p}^{n+1} \mid \mathbf{p}^n)$ is strictly larger than $g(\mathbf{p}^n \mid \mathbf{p}^n)$ unless $\mathbf{p}^{n+1} = \mathbf{p}^n$. Because $g(\mathbf{p} \mid \mathbf{p}^n)$ minorizes $L(\mathbf{p})$, this ascent property carries over to $L(\mathbf{p})$. With minor notational changes, the same arguments apply to the directed graph model.

Log P-Value Approximations

Since the extreme right-tail probabilities of the Poisson distribution lead to computer underflows, we must resort to approximation. Let the Poisson random deviate X have mean λ. For n much larger than λ, we find that

$$\begin{aligned}
\Pr(X \geq n) &= \sum_{k=n}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} \\
&= \frac{e^{-\lambda}\lambda^n}{n!} \sum_{k=0}^{\infty} \frac{\lambda^k n!}{(n+k)!} \\
&\leq \frac{e^{-\lambda}\lambda^n}{n!} \sum_{k=0}^{\infty} \left(\frac{\lambda}{n}\right)^k \\
&= \frac{e^{-\lambda}\lambda^n}{n!} \left(\frac{1}{1-\frac{\lambda}{n}}\right) \\
&= \frac{e^{-\lambda}\lambda^n}{(n-1)!} \frac{1}{n-\lambda}.
\end{aligned}$$

Because n is large, we can approximate $(n-1)!$ by Stirling's formula

$$(n-1)! \approx \sqrt{2\pi} n^{n-1/2} e^{-n}.$$

This allows us to take logarithms of $\Pr(X \geq n) \approx \frac{e^{n-\lambda}\lambda^n}{\sqrt{2\pi}n^{n-1/2}(n-\lambda)}$ in the construction of our tables.

REFERENCES

Hiriart-Urruty J-B (2004) *Fundamentals of Convex Analysis*. Springer-Verlag, New York.

Lange K (2004) *Optimization*. Springer-Verlag, New York.

APPENDIX TABLES AND FIGURES

Table 1. Convergence results for each of the 5 real datasets. Note that convergence was defined as a change in loglikelihood of less than 10^{-8} percent of the previous loglikelihood. Time given is for a dual processor computer running at 2.4 GHz. Time is given in seconds (s).

Dataset	# Nodes	# Edges	# Iterations	Time(s)
Letter Pairs	27	503,951	21	42
C. Elegans	281	6,417	23	9
Protein Ints.	9,213	88,456	18	741
Word Pairs	10,789	137,338	24	1,415
Rad. Hybrid	20,145	825,551,643	29	14,903

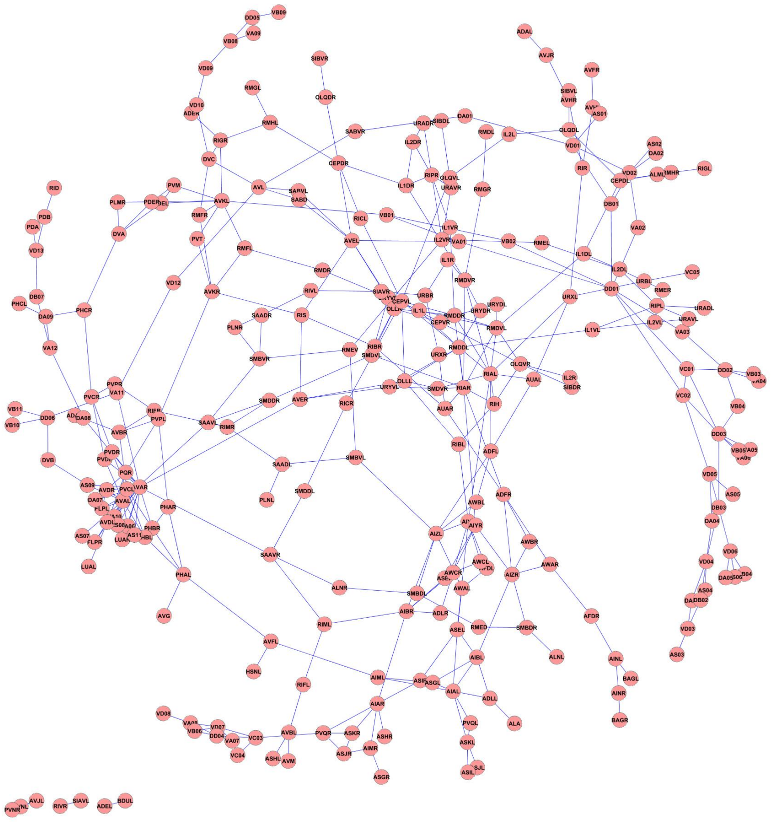

Fig. 1. Graph of *C. Elegans* neural network with a p-value of 10^{-6}.

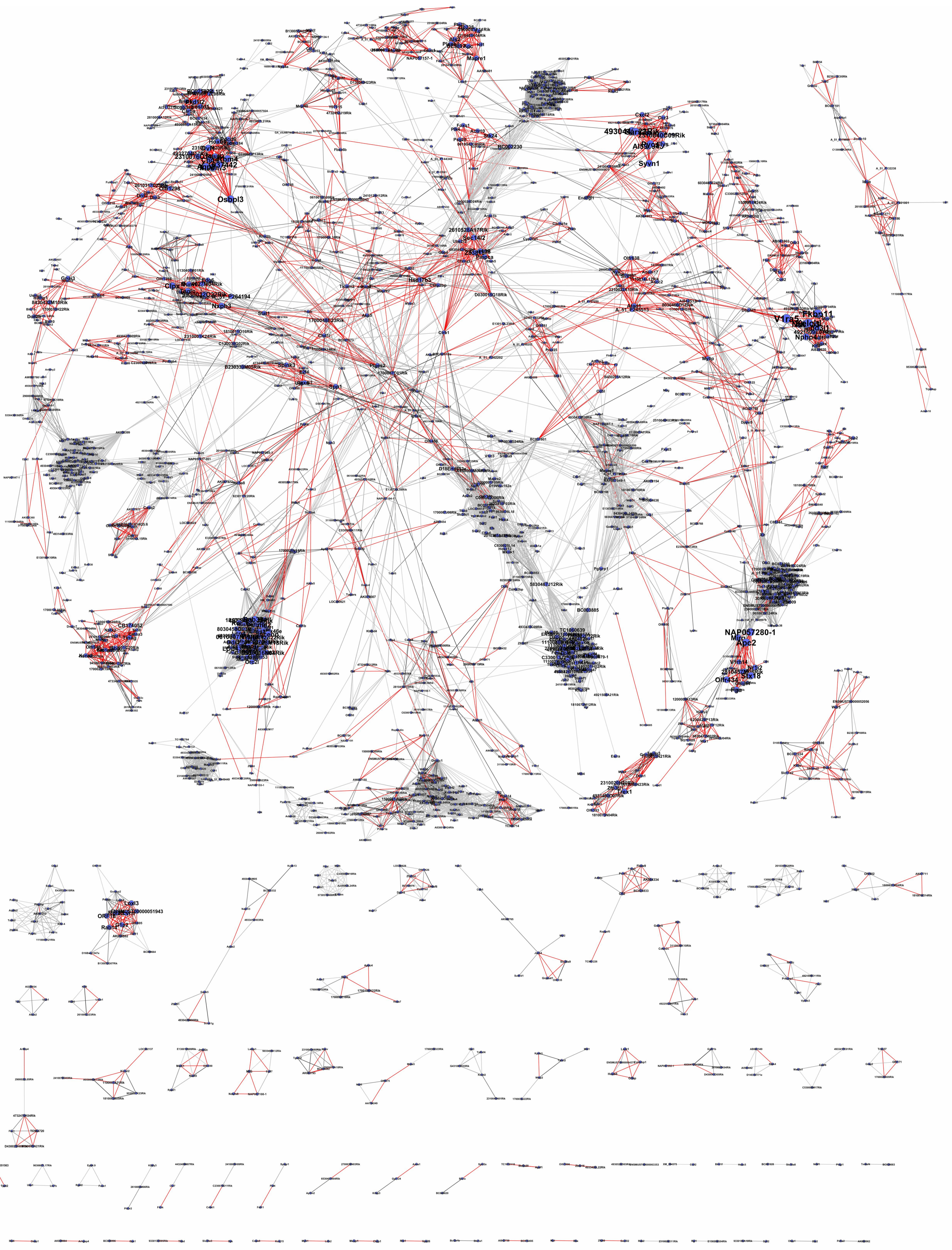

Fig. 2. Graph of the Radiation Hybrid network. In this graph, node size is proportional to a node's estimated propensity. Also, the darker the edge, the more significant the connection; red lines highlight the most significant connections.