Googling the Brain: Discovering Hierarchical and Asymmetric Network Structures, with Applications in Neuroscience

Jonathan J. Crofts and Desmond J. Higham

Abstract. Hierarchical organization is a common feature of many directed networks arising in nature and technology. For example, a well-defined message-passing framework based on managerial status typically exists in a business organization. However, in many real-world networks, such patterns of hierarchy are unlikely to be quite so transparent. Due to the nature in which empirical data are collated, the nodes will often be ordered so as to obscure any underlying structure. In addition, the possibility of even a small number of links violating any overall "chain of command" makes the determination of such structures extremely challenging.

Here we address the issue of how to reorder a directed network to reveal this type of hierarchy. In doing so, we also look at the task of quantifying the level of hierarchy, given a particular node ordering. We look at a variety of approaches. Using ideas from the graph Laplacian literature, we show that a relevant discrete optimization problem leads to a natural hierarchical node ranking. We also show that this ranking arises via a maximum likelihood problem associated with a new range-dependent hierarchical random-graph model. This random-graph insight allows us to compute a likelihood ratio that quantifies the overall tendency for a given network to be hierarchical. We also develop a generalization of this node-ordering algorithm based on the combinatorics of directed walks. In passing, we note that Google's PageRank algorithm tackles a closely

related problem, and may also be motivated from a combinatoric, walk-counting view-point. We illustrate the performance of the resulting algorithms on synthetic network data, and on a real-world network from neuroscience where results may be validated biologically.

1. Introduction

Determining hidden structure and substructure within a complex network provides us with a host of clues concerning both its functional capabilities and its evolutionary history. This motivates the need for quantitative tools to discover significant topological features, such as well-connected communities, bipartite structures, bottlenecks, motifs, hubs, and authorities [Costa et al. 07, Fortunato 10, Girvan and Newman 02, Holme et al. 03, Sporns and Kotter 04]. In this work, we focus on a particular type of structure, namely that of a directed hierarchy, a notion that will be defined more precisely as we proceed.

We begin by noting that the term *hierarchical* may be used to denote a fractal type of network organization, whereby smaller modules repeat in a self-similar manner [Ravasz and Barabási 03, Zhou et al. 06]. In this work we are using the alternative meaning of hierarchical to denote a directed network structure that supports a well-defined message-passing or "chain of command" scenario such that the processing of information or the exercise of managerial control proceeds sequentially in a top-to-bottom fashion. This is the viewpoint taken in the recent articles [Chen et al. 07, Muchnik et al. 07, Trusina et al. 04]. Our work differs in that we aim to use first-principle arguments based on quantitative measures of network hierarchy in order to derive algorithms that reveal these structures. We also aim to provide a quantitative summary of the overall tendency of a given network to be hierarchical, and in doing so, we introduce a new class of random graphs with a hierarchical structure.

We use the following notation. Given an unweighted directed network consisting of N nodes, we denote by $A \in \mathbb{R}^{N \times N}$ the corresponding adjacency matrix. So A is generally asymmetric and has $a_{ij} = 1$ if there is an edge from node i to node j, and $a_{ij} = 0$ otherwise. We assume $a_{ii} = 0$ for $i = 1, \ldots, N$, discounting self-loops. The out- and in-degrees of node k are specified as

$$\deg_k^{\mathrm{out}} := \sum_j a_{kj} \quad \text{and} \quad \deg_k^{\mathrm{in}} := \sum_i a_{ik},$$

respectively. We denote by $\mathcal{P}$ the set of all permutations of the integers $1, 2, \ldots, N$, and use p_i to denote the ith component of member $p \in \mathcal{P}$. The

Euclidean vector norm is denoted by $\| \cdot \|_2$, and $\mathbf{1} \in \mathbb{R}^N$ represents the vector with all components equal to one.

The layout of the paper is as follows. In Section 2 we recall how the widely used Fiedler vector is relevant as the solution to an optimization problem that emphasizes similarity between nodes. We then introduce a new optimization problem as a means to quantify hierarchy, and show how it can be solved. This results in a simple network-reordering algorithm based on the difference between out- and in-degrees.

In Section 3, we exploit ideas from random-graph theory that can be used to reorder a network from a maximum-likelihood perspective. We show that the algorithm from Section 2 may be viewed as a maximum-likelihood reordering under a new, directed, range-dependent random-graph hypothesis. In addition to providing an alternative motivation for the algorithm, this viewpoint also allows us to quantify the level of hierarchy in a network by comparing log-likelihood ratios for hierarchical versus nonhierarchical range-dependent structures. We give some illustrative computational results on synthetically generated networks, quantifying the extent to which this approach is tolerant of noise.

Section 4 then looks at an alternative, combinatoric, approach. We show how the level of hierarchy in a particular node can be conveniently quantified by counting directed walks. We show in Section 5 how Google's PageRank algorithm performs a related task. In Section 6 we evaluate the new methods on a neuronal network whose hierarchical structure has biological significance.

2.   Out-Degree Minus In-Degree

To motivate the work in this section, we begin with the *two-sum* [Barnard et al. 95, Higham 03]

$$\sum_{i=1}^{N}\sum_{j=1}^{N}(i-j)^2 a_{ij}.$$

This nonnegative quantity is small when the presence of edges is biased toward nodes that have nearby indices. Equivalently, it is small when the nonzeros in the adjacency matrix appear close to the diagonal. Reordering the nodes corresponds to the mapping $i \to p_i$ for some permutation $p \in \mathcal{P}$, and the task of finding a minimum two-sum reordering may thus be written

$$\min_{p \in \mathcal{P}} \sum_{i=1}^{N}\sum_{j=1}^{N}(p_i - p_j)^2 a_{ij}. \tag{2.1}$$

In general, this discrete optimization problem is computationally intractable, and it is therefore common to consider a relaxed version in which p is allowed to take real values. To avoid redundancies from scaling and shifting, we also impose constraints $\|p\|_2 = 1$ and $p^T \mathbf{1} = 0$. This leads to

$$\min_{\substack{p \in \mathbb{R}^N \\ \|p\|_2 = 1 \\ p^T \mathbf{1} = 0}} \sum_{i=1}^{N} \sum_{j=1}^{N} (p_i - p_j)^2 a_{ij}. \tag{2.2}$$

We may now introduce the graph Laplacian $L := D - B$, where $B := \frac{1}{2}(A + A^T)$ and the diagonal matrix D has $d_{ii} = \sum_k b_{ik}$. In the case that the graph represented by B is connected, L is symmetric positive semidefinite with a single eigenvalue equal to zero and corresponding eigenvector proportional to $\mathbf{1}$. It follows that (2.2) is solved by taking p to be the *Fiedler vector* v, that is, the eigenvector corresponding to the second-smallest eigenvalue of L. We refer to [Chung 97, Strang 08, Van Driessche and Roose 95] for further details, noting that most authors treat the case in which A is symmetric.

Having obtained a relaxed solution v, we may recover a permutation $p \in \mathcal{P}$ by the natural procedure of ordering the nodes according to their real-valued components. More precisely, we compute $p \in \mathcal{P}$ such that

$$v_i \leq v_j \iff p_i \leq p_j, \tag{2.3}$$

with some rule for treating ties.

To illustrate this idea, the upper left picture in Figure 1 shows the adjacency matrix for a network showing a strong preference for short-range edges (to produce this matrix, we computed an instance of Grindrod's range-dependent random-graph model described in Section 3, with $\beta = 0.015$ and $N = 100$). After applying a random shuffle to the nodes, we obtained the adjacency matrix shown in the upper middle picture in Figure 1; any evidence of the range-dependent structure has been completely destroyed. Finally, the upper right picture in Figure 1 shows the adjacency matrix when reordered via the Fiedler vector. We see that the hidden structure has been recovered. The respective values of the two-sum going from left to right are 32037, 1283966, and 30155. As mentioned above, the reordering from the Fiedler vector is not guaranteed to minimize the two-sum, since it solves only a relaxed version of the problem. However, in this case we note that it gives a two-sum that is below the value produced by the original lattice ordering.

We now consider how this approach can be adapted to discover hierarchical structure. We use the convention that nodes are to be ordered in descending order of importance, so node 1 is at the top of the hierarchy and node N at the
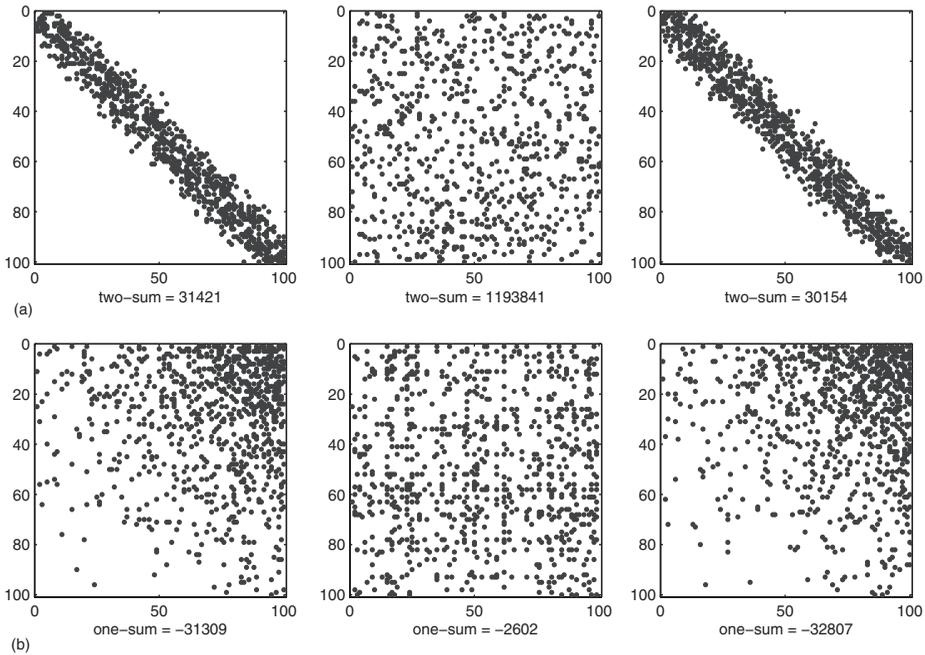
Figure 1. Left: Adjacency matrix for a network with (a) short-range connectivity structure and (b) hierarchical connectivity structure. Middle: In both cases, the connectivity structure is hidden via a random shuffle of the network nodes. Right: The respective structures are recovered via (a) a Fiedler vector reordering and (b) reordering based on out-degree minus in-degree (see (2.6)).

bottom. For a given network and a given node ordering, we may then introduce the new concept of a *directed one-sum*

$$\sum_{i=1}^{N}\sum_{j=1}^{N}(i-j)a_{ij} \tag{2.4}$$

to quantify the level of hierarchy. In this expression, each edge is involved in the overall sum. An edge that respects the hierarchy, $i < j$, contributes a negative amount, with the weight $i - j$ being more negative when i is further up the hierarchy than j. Similarly, an edge that violates the hierarchy, $i > j$, contributes a positive amount, with the weight $i - j$ being more positive when i is further down the hierarchy than j. It follows that a more negative value for the directed one-sum corresponds to a more hierarchical combination of network and node ordering.

In order to search for the best way to reorder a given network in a manner that reveals hierarchical structure, it is therefore reasonable to minimize the directed one-sum:

$$\min_{p \in \mathcal{P}} \sum_{i=1}^{N} \sum_{j=1}^{N} (p_i - p_j) a_{ij}. \tag{2.5}$$

This is very simply rewritten as

$$\min_{p \in \mathcal{P}} \sum_{i=1}^{N} p_i \left(\deg_i^{\text{out}} - \deg_i^{\text{in}} \right),$$

and it follows that the problem (2.5) is solved by ordering the nodes according to the difference between out- and in-degrees:

$$\deg_i^{\text{out}} - \deg_i^{\text{in}} \geq \deg_j^{\text{out}} - \deg_j^{\text{in}} \iff p_i \leq p_j. \tag{2.6}$$

We emphasize that in contrast to the two-sum case (2.1)–(2.3), this ordering solves the discrete formulation of the problem, not just a relaxed version of it.

To illustrate this idea, the lower left picture in Figure 1 shows the adjacency matrix for a network showing a strong hierarchical structure (to produce this matrix, we computed an instance of the directed range-dependent random-graph model described in Section 3, with $\alpha = 0.025$ in (3.2) and $N = 100$). An arbitrary node shuffle produced the adjacency matrix shown in the lower middle picture in Figure 1. Finally, the lower right picture in Figure 1 shows the adjacency matrix when reordered via the out-degree minus in-degree (2.6). We see that the hidden structure has been recovered. The respective one-sum values from left to right are -31309, -2602, and -32807, and we know from the derivation above that the latter value is the global minimum.

We note that the authors in [Muchnik et al. 07] considered the *ratio* of in-degree to out-degree in order to rank a node within a hierarchy, without giving any justification for the measure. Of course, this is equivalent to a ranking based on the log-difference $\log(\deg_i^{\text{out}}) - \log(\deg_i^{\text{in}})$, whereas the ranking that we have derived via the directed one-sum does not use logs. Additionally, the new ranking derived here is also able to deal with nodes containing no outgoing connections.

We also note a similarity between directed one-sum minimization and the task of finding a *topological ordering* of a directed graph. Such an ordering may be viewed as having no positive values in the directed one-sum (2.4). Alternatively, it may be viewed as an ordering in which all nonzeros in the adjacency matrix have been placed in the upper triangle. This is possible if and only if we have a directed acyclic graph (DAG) [Bang-Jensen and Gutin 02]. The minimization problem (2.5) has a different emphasis in that (a) we seek a complete ordering of

the nodes rather than a partial ordering, and (b) we wish to apply the concept to general directed networks rather than DAGs. Further, in (2.5), we measure the extent to which the hierarchy has been respected or violated in the sense that a link from node 1 to node 100 makes a more favorable contribution than a link from node 100 to node 99, and similarly, a link from node 100 to node 1 makes a more unfavorable contribution than a link from node 100 to node 99. However, in the spirit of topological ordering for DAGs, in Section 6 we use the percentage of nonzeros in the upper triangle of the reordered adjacency matrix as one way to compare algorithms.

3. Random-Graph Viewpoint

In this section we show that the out-minus-in ordering (2.6) can also be justified via a random-graph argument, and use this as a means to quantify the amount of hierarchy in a given network.

The general class of range-dependent random graphs introduced in [Grindrod 02, Grindrod 03] may be defined as follows.

Definition 3.1. For a given function g that maps from $\{1, 2, \ldots, N-1\}$ to $[0, 1]$, a *range-dependent random graph* (RDRG) is an undirected graph that has an edge from node i to node j with independent probability $g(|i - j|)$.

With this definition, we may think of node i as being positioned at location i on the integer lattice. The chance of two nodes being linked is then a function of their lattice distance $|i - j|$.

The upper left picture of Figure 1 shows an instance of an RDRG with g taking the functional form of (3.4), with $\beta = 0.15$. Because g is a decaying a function, "long-range" connections are less likely than those at "short range."

We now introduce a natural analogue of Grindrod's RDRG class of models that allows directed and hierarchical networks to be generated at random.

Definition 3.2. For a given function f that maps from $\{1, 2, \ldots, 2N-1\}$ to $[0, 1]$, a *directed range-dependent random graph* (dRDRG) has an edge from node i to node j with independent probability $f(i - j + N)$.

A hierarchical structure will arise when f is monotonically decreasing in this dRDRG setting; in that case, the edge from node $i = 1$ to node $j = N$ is the most likely, and the reverse edge from $i = N$ to $j = 1$ is the least likely.

Given a network and an ordering, the likelihood that these data came from the dRDRG model is given by

$$\prod_{a_{ij}=1} f(i - j + N) \times \prod_{a_{ij}=0} (1 - f(i - j + N)).$$

We may then ask for the reordering that is most likely under this model:

$$\max_{p \in \mathcal{P}} \prod_{a_{ij}=1} f(p_i - p_j + N) \times \prod_{a_{ij}=0} (1 - f(p_i - p_j + N)). \qquad (3.1)$$

The next result connects this problem to the directed one-sum minimization.

Proposition 3.3. *The one-sum minimization problem* (2.5) *is equivalent to the maximum likelihood reordering problem* (3.1) *when the directed range-dependency takes the form*

$$f(i - j + N) = \frac{e^{-\alpha(i-j+N)}}{1 + e^{-\alpha(i-j+N)}}, \qquad (3.2)$$

for any fixed $\alpha > 0$.

Proof. We may follow the technique used in [Grindrod 02] in the undirected case by rewriting (3.1) as

$$\max_{p \in \mathcal{P}} \prod_{a_{ij}=1} \frac{f(p_i - p_j + N)}{1 - f(p_i - p_j + N)} \times \prod_{\text{all } i,j} (1 - f(p_i - p_j + N)).$$

The second factor, the probability of a null graph, is independent of p, so an equivalent problem is

$$\max_{p \in \mathcal{P}} \prod_{a_{ij}=1} \frac{f(p_i - p_j + N)}{1 - f(p_i - p_j + N)}.$$

When f has the form in (3.2), this may be rewritten as

$$\max_{p \in \mathcal{P}} \prod_{a_{ij}=1} e^{-\alpha(p_i - p_j + N)}.$$

Taking logs and negating the objective function leads to

$$\min_{p \in \mathcal{P}} \alpha \sum_{i=1}^{N} \sum_{j=1}^{N} (p_i - p_j + N) \, a_{ij},$$

which is clearly equivalent to (2.5). ☐

The picture on the left in Figure 2 shows an instance of the hierarchical dR-DRG model, with the nodes ordered according to the range-dependency, in the
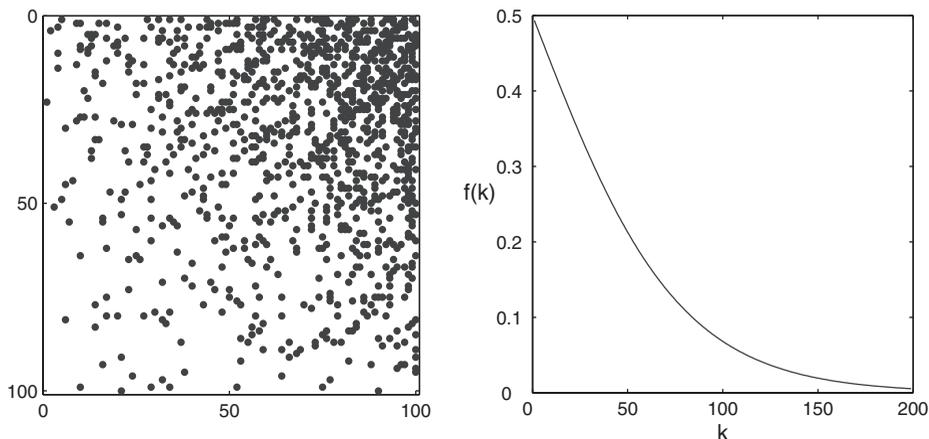
Figure 2. An instance of a hierarchical dRDRG with $N = 100$ and $\alpha = 0.0262$ (left). The model connects node i to node j with probability $f(i - j + N)$, where f has the form shown on the right.

case that $N = 100$ and $\alpha = 0.0262$ in (3.2). We see that the structure is consistent with the aims of the directed one-sum minimization: Nonzeros are most prevalent at the upper right-hand corner of the matrix, and the density decreases as we move toward the lower left-hand corner. In the right of the figure we plot the range-dependency function f from (3.2).

We see from Definition 3.1 that with Grindrod's RDRG model, we have, for a given decay function g, $a_{ij} = 1$ with probability $g(|i - j|)$ and $a_{ij} = 0$ otherwise. It was shown in [Higham 03] that the corresponding maximum likelihood reordering problem

$$\max_{p \in \mathcal{P}} \prod_{a_{ij}=1} g(|p_i - p_j|) \times \prod_{a_{ij}=0} (1 - g(|p_i - p_j|)) \qquad (3.3)$$

is equivalent to two-sum minimization (2.1) for the particular range-dependency

$$g(|i - j|) = \frac{e^{-\beta(i-j)^2}}{1 + e^{-\beta(i-j)^2}}, \qquad (3.4)$$

where we have introduced a free parameter $\beta > 0$ to be compatible with the general case specified in Proposition 3.3.

This insight puts us in a position to compare the likelihood that a given network arose from each of these two model classes. We may summarize this as Algorithm 1.

As a proof of principle, we give results for a large-scale experiment using synthetically generated networks. In each test, we began with an instance of

---

Algorithm 1 Hierarchical reordering and network comparison.

---

1. Fit the parameters α and β in the two models. This can be done by matching the number of edges in the given network to the expected number of edges in the random-graph model. The latter is a monotonic function of the free parameter in both cases, so a simple bisection root-finding method is guaranteed to succeed.

2. Compute the maximum likelihood from the hierarchical model, say $\mathcal{L}_{\mathrm{h}}$, which is given by inserting the minimum directed one-sum ordering into the objective function in (3.1) with f defined in (3.2). Compute the approximate maximum likelihood from the nonhierarchical model, say $\mathcal{L}_{\mathrm{nh}}$, which is given by inserting the Fiedler vector ordering into the objective function in (3.3) with g defined in (3.4). (This is approximate because the Fiedler vector solves a relaxed version of the minimum two-sum problem.)

3. Compute the normalized log likelihood ratio

$$L := \frac{2}{N(N-1)} \log\left(\frac{\mathcal{L}_{\mathrm{nh}}}{\mathcal{L}_{\mathrm{h}}}\right). \qquad (3.5)$$

A negative ratio indicates that the network data support the hierarchical range-dependency over the nonhierarchical alternative.

---

a network from one of the two random classes under consideration. To assess robustness to noise, we then added spurious links at random. For each network instance, we placed a new link between all possible pairs of nodes with small independent probability q. If the likelihood ratio correctly identified the structure of interest, we increased the noise parameter q and repeated the analysis. Tables 1 and 2 record the smallest value of q (averaged over an ensemble of 1000 random graphs) for which the likelihood ratio incorrectly classified the structure. We used graphs of order $N = 50, 100, 200, 500, 1000$, which had on average $(N^2 - N) \times s$ edges; here $s \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ denotes the proportion of edges present, on average, for each network instance.

From Tables 1 and 2, we see that the likelihood approach continues to classify the network models correctly even after some 10%–15% of additional links have been added. Note that robustness generally improves as we increase the proportion of connections present in the initial instance, a trend that is particularly evident for the hierarchical model. This observation is consistent with the idea that the presence of additional links translates into additional information concerning network architecture. To match the case of typical real networks, we have focused on sparse structures. If we tuned α and β to allow for increasingly dense

			N		
s	50	100	200	500	1000
0.1	0.1150	0.1147	0.1149	0.1122	0.1151
0.2	0.1360	0.1354	0.1370	0.1361	0.1359
0.3	0.1472	0.1472	0.1479	0.1471	0.1478
0.4	0.1548	0.1552	0.1550	0.1547	0.1546
0.5	0.1572	0.1573	0.1577	0.1571	0.1572

Table 1. Proportion of spurious links tolerated in discovering a hierarchical dR-DRG network.

connectivity patterns, then intuitively, we would expect to see the test become less reliable: As both network models approach the complete graph, they become indistinguishable.

4.   Directed Walks and Matrix Functions

The out-minus-in ordering studied in the previous two sections is computationally convenient and was justified from first principles. However, it has the possible drawback that because it is based on integer-valued quantities, nodes may frequently be tied. In particular, for the simple case of a directed binary tree, all nodes except the root and leaves have an equal out-minus-in score.

			N		
s	50	100	200	500	1000
0.1	0.1290	0.1259	0.1256	0.1271	0.1274
0.2	0.1356	0.1339	0.1352	0.1353	0.1348
0.3	0.1366	0.1360	0.1368	0.1363	0.1357
0.4	0.1362	0.1362	0.1360	0.1365	0.1359
0.5	0.1359	0.1358	0.1358	0.1357	0.1355

Table 2. Proportion of spurious links tolerated in discovering a nonhierarchical RDRG network.

A more global strategy that looks beyond the degree structure was considered in [Muchnik et al. 07], where the concept of *attraction basin hierarchy* was introduced. The resulting algorithm is based on counting the number of shortest paths to/from a particular node from/to all other nodes, which is a hard combinatorial problem in general. Here, we propose an alternative measure that uses directed walks rather than paths.[1] Recent work has shown that walk-based measures provide an effective tool for determining other types of connectivity patterns [Crofts et al. 10, Estrada and Hatano 08, Estrada et al. 08, Grindrod 11], and there are at least four specific arguments in their favor:

- information does not generally flow along geodesics [Borgatti 05, Newman 05];

- walk counts are less sensitive than path-length counts to spurious or missing edges [Grindrod 11];

- walk counts can be computed conveniently using basic operations in linear algebra [Estrada and Higham 10];

- as we will show below, Google's successful PageRank algorithm has a walk-based interpretation.

The computational convenience arises because the element $(A^n)_{ij}$ counts the number of directed walks of length n that start at node i and terminate at node j. We may then introduce coefficients $c_0, c_1, c_2, c_3, \ldots$ and consider the expansion

$$F(A) = c_0 I + c_1 A + c_2 A^2 + c_3 A^3 + \cdots . \tag{4.1}$$

Because longer walks are generally (a) more numerous and (b) less important than shorter walks, it is reasonable to choose a decreasing sequence of coefficients. It was suggested in [Estrada and Rodríguez-Velázquez 05] that $c_k = 1/(k!)$, so that $F(A) = \exp(A)$, and this choice has proved to be very useful in many scenarios [Crofts et al. 10, Crofts and Higham 09, Estrada and Hatano 07, Estrada and Hatano 08, Estrada et al. 08, Estrada et al. 09]. The resolvent function $F(A) = (I - \delta A)^{-1}$, corresponding to $c_k = \delta^k$, was considered in [Estrada and Higham 10] and can be traced back to earlier work on node centrality indices in social network analysis [Katz 53, Newman 10]. It is also possible to interpret δ as the probability that a message will successfully traverse an edge.

---

[1] A walk differs from a path in that nodes and edges may be reused during the traversal.

We may now classify the hierarchical rank of a node by measuring

- $\sum_{j \neq i} (F(A))_{ij}$, which quantifies how effectively node i is able to pass information to or exert control over the other nodes in the network;

- $\sum_{j \neq i} (F(A))_{ji}$, which quantifies how effectively the other nodes in the network are able to pass information to or exert control over node i.

We therefore propose that node i be assigned the hierarchical ranking

$$r_i = \sum_{j \neq i} (F(A))_{ij} - \sum_{j \neq i} (F(A))_{ji}, \qquad (4.2)$$

leading to a reordering in descending rank order, that is,

$$r_i \geq r_j \iff p_i \leq p_j.$$

Note that the walk-based hierarchy measure of (4.2) can be considered a natural generalization of the degree-based algorithm introduced earlier, in the sense that by choosing $F(A) = A$, we recover the out-degree minus in-degree ordering.

5.   Links to Google's PageRank

The well-known PageRank algorithm is used by Google to quantify the importance of web pages based on the hyperlink topology of the World Wide Web (WWW) [Langville and Meyer 06, Page et al. 98]. Letting A represent the web adjacency matrix, so that $a_{ij} = 1$ means that page i has a hyperlink to page j, this algorithm assigns rank according to the vector

$$\left(I - \theta A^T D^{\text{out}\,-1} \right)^{-1} \mathbf{1}, \qquad (5.1)$$

with a higher value in component i denoting more importance for page i. Here $\theta \in (0,1)$ is a free parameter and D^{out} is the diagonal out-degree matrix, so $D_{ii}^{\text{out}} = \text{deg}_i^{\text{out}}$. We assume for the moment that $\text{deg}_i^{\text{out}} \geq 1$ for all i. Writing $\widehat{A} := A^T D^{\text{out}\,-1}$, we see that for sufficiently small θ, the ranking in (5.1) may be expanded as

$$\left(I + \theta \widehat{A} + \theta^2 \widehat{A}^2 + \theta^3 \widehat{A}^3 + \cdots \right) \mathbf{1}.$$

This shows that PageRank may be interpreted as a directed-walk counting algorithm, with a few twists:[2]

---

[2] We note that this is distinct from the well-known *random walk* interpretation of PageRank [Langville and Meyer 06, Page et al. 98].

Link direction is reversed and edges are scaled. The new matrix $\widehat{A}$ has $\widehat{a}_{ij} \neq 0$ if and only if there is a hyperlink from page j to page i. This is intuitively reasonable—the $j \mapsto i$ hyperlink may be interpreted as page j deferring to, or handing control over to, page i. (In the WWW context it is also extremely pertinent that page i has no direct influence over the creation of a $j \mapsto i$ hyperlink. This makes it difficult to boost artificially your own page's ranking.) Supposing that the hyperlinks $i_{k+1} \mapsto i_k$, $i_k \mapsto i_{k-1}, \ldots, i_2 \mapsto i_1$ exist on the WWW, then the walk of length k given by $i_1 \to i_2 \to \cdots \to i_{k+1}$ exists in the reversed network, but in addition to the θ^k scaling that penalizes long walks, there is also a scaling

$$\frac{1}{\deg_2^{\mathrm{out}}} \frac{1}{\deg_3^{\mathrm{out}}} \cdots \frac{1}{\deg_{k+1}^{\mathrm{out}}}$$

penalizing walks involving "promiscuous" nodes that make themselves available for many other such walks. This method for nullifying the influence of overactive nodes that are likely to be lowly ranked is clearly relevant in the WWW setting. However, in the context of this work, where we are seeking to discover a hierarchical "chain of command," it seems less appropriate to normalize in this way. If node i gives orders to (i.e., has a hyperlink from) a node that also gets orders from (i.e., has hyperlinks to) many other nodes, then this could be regarded as sound evidence for placing node i near the top of the hierarchy.

Closed walks are considered. Node i is given a ranking based on the sum over all j of a weighted count of directed walks that begin at i and end at j, including the closed-walk case $j = i$.

6. Local *C. elegans* Network

In the PhD thesis [Durbin 87], the author went about the task of sorting the neurons in the nerve ring of *C. elegans*[3] vertically in such a way that as many of the synapses as possible pointed downward. Durbin used an ad hoc combinatoric algorithm to create an ordering. Here, we propose to repeat Durbin's analysis using the automated algorithms that we have derived. We consider a local subnetwork of 131 frontal neurons and 782 chemical synapses of the neuronal network for *C. elegans* [Kaiser and Hilgetag 06, Pan et al. 10, Varshney

---

[3] A 1-mm-long transparent roundworm.
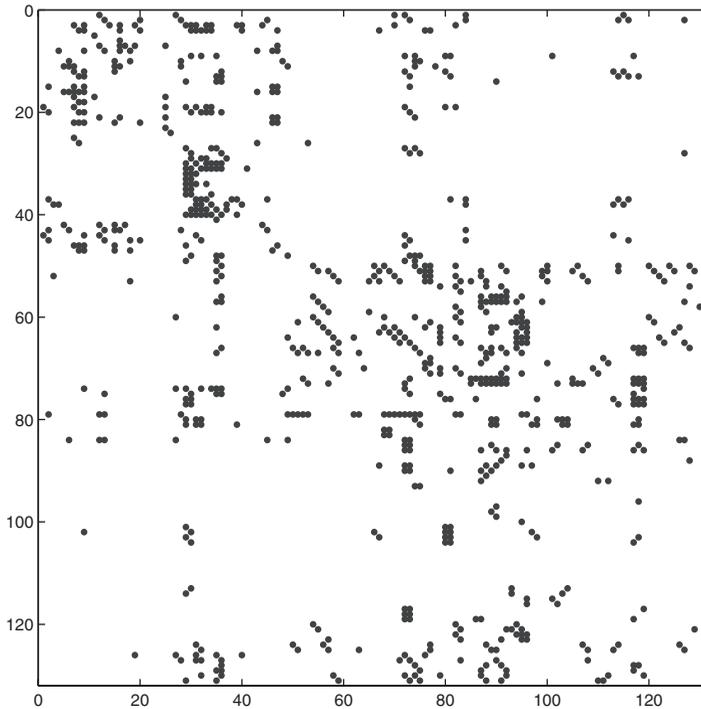
Figure 3.   Adjacency matrix for the local network of 131 frontal neurons for *C. elegans*.

et al. 11]. We exclude gap junctions[4] from the analysis, since (a) it remains an open question whether such connections exhibit directionality, and (b) current experimental techniques are unable to extract this information in any case. The adjacency matrix is shown in Figure 3.

Figure 4 shows the results of reordering the local *C. elegans* network based on

  (i)   out-degree minus in-degree,

 (ii)   the walk-based measure (4.2) with $F(A) = \exp(A)$,

(iii)   the walk-based measure (4.2) with $F(A) = (I - \delta A)^{-1}$ and $\delta = 0.025$,

(iv)   Google's PageRank.

---

[4] Gap junctions are channels that provide electrical coupling between neurons.
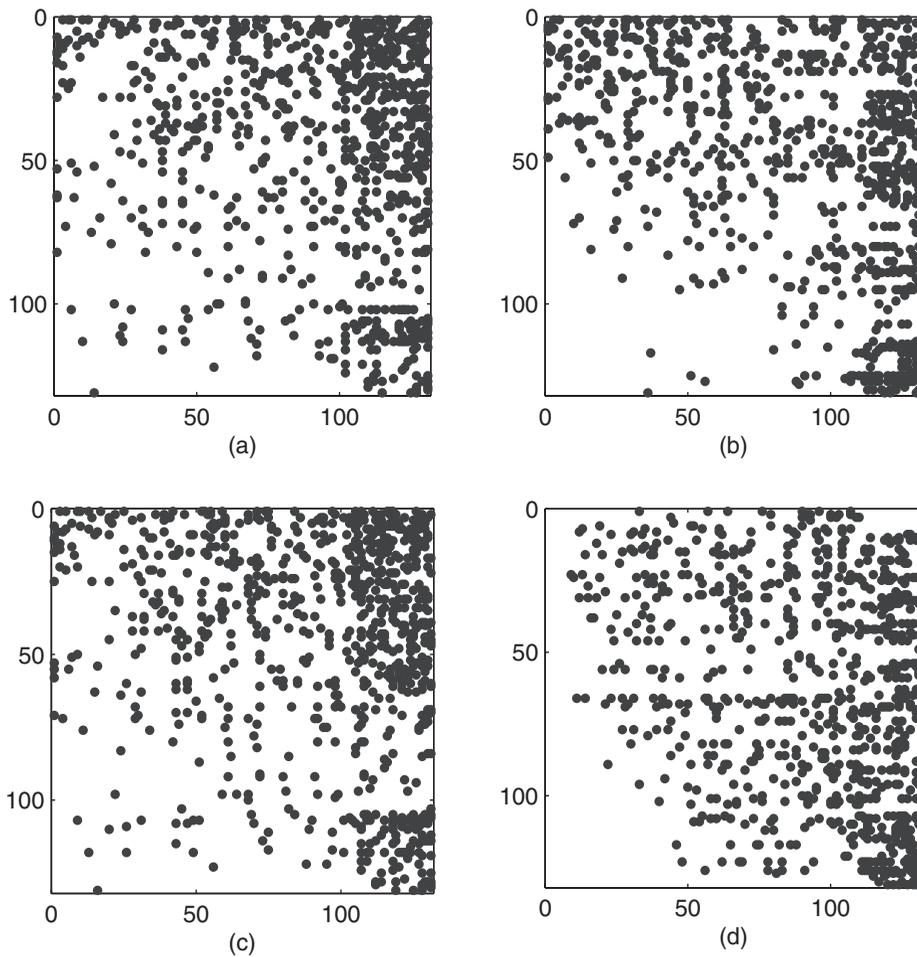
Figure 4. Results for the *C. elegans* neuronal network: (a) reordered using out-degree minus in-degree; (b) reordered using the walk-based measure with $F(A) = \exp(A)$; (c) reordered using the walk-based measure with $F(A) = (I - \delta A)^{-1}$ and $\delta = 0.025$; (d) reordered using Google's PageRank algorithm.

In the latter case, dangling nodes, for which $\deg_i^{\mathrm{out}} = 0$, require a slight modification of the Google matrix in (5.1) [Langville and Meyer 06]. We used the widely quoted teleporting parameter value $\theta = 0.85$.

Table 3 quantifies the ability of each method to discover hierarchical structure. The first row gives the directed one-sum, that is, the objective function in (2.5), for each ordering. We note from Proposition 3.3 that the out-minus-in ordering

	out minus in	exponential	resolvent ($\delta = 0.025$)	PageRank
Directed one-sum	-32857	-27198	-32809	-22649
Proportion of nonzeros in upper triangle	0.8095	0.7724	0.8056	0.7366

Table 3. Hierarchy scores for the *C. elegans* network.

minimizes this quantity. From Table 3, we see that the directed one-sum in (2.4) attains a minimum value of -32857. Next came the walk-based algorithms, with scores of -32809 using the resolvent function and -27198 using the matrix exponential. Finally, the ordering obtained via the PageRank algorithm performed poorest according to this measure, with a score of -22649.

The second row in Table 3 records the proportion of nonzeros in the upper triangle of the adjacency matrix. With this alternative measure, there is no extra benefit/penalty from violating/exploiting the hierarchy with a "long-range" link that spans many intermediate nodes. Instead, we simply count the proportion of links that respect the hierarchy. We see from the table that the out-minus-in ordering is marginally better than that given by the walk-based algorithms according to this measure, with all methods scoring between 77% to 81%, while PageRank achieved around 73%.

To justify the choice of $\delta = 0.025$ in the resolvent-based reorderings, Figure 5 shows how the proportion of nonzeros in the upper triangle (left) and directed one-sum (right) vary as a function of δ. Both measures start to degrade beyond this level.

The normalized log-likelihood ratio in (3.5) was $L = -1.9319$, giving further support for the visually compelling evidence in Figure 4 that this network has a strong hierarchical element.

Let us now focus on the biological significance of these results in the particular case of the walk-based measure with $F(A) = (I - \delta A)^{-1}$ and $\delta = 0.025$. Table 4 summarizes the results for the reordered *C. elegans* data, where we report those neuronal classes represented by nodes in the top and bottom 10% of the reordered network. Perhaps most noteworthy is the fact that neuronal classes representing sensory neurons are highly prevalent at the top of the hierarchy (some 85%), while the foot of the hierarchy consists wholly of a mixture of motor neurons and command interneurons. In general, we found that the ordering returned was
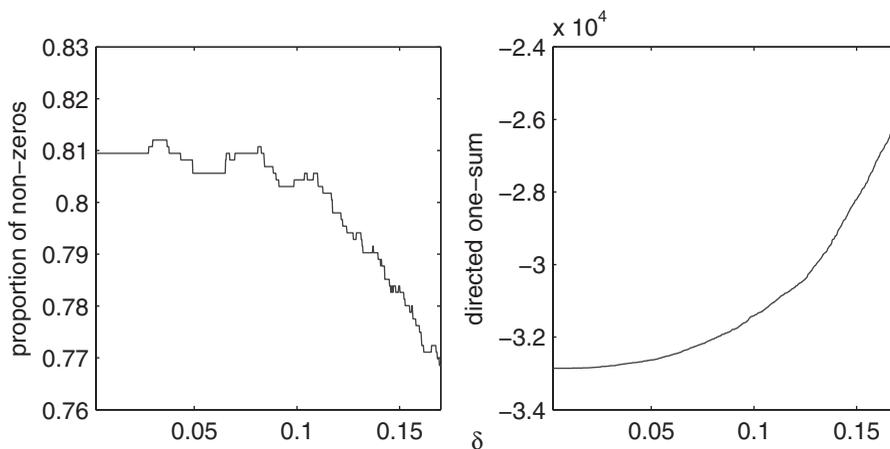
Figure 5. Plots of both hierarchy scores versus δ using the resolvent function $F(A) = (I - \delta A)^{-1}$ to reorder the *C. elegans* network.

in good agreement with that of Durbin, with sensory neurons placed at the top, motor neurons at the bottom, and interneurons in between.

The command interneuron classes AVA, AVB, AVD, and AVE appearing at the bottom of the hierarchy in Table 4 have been identified in previous studies [Morita et al. 01] as being highly connected to both sensory

	Top 10%		Bottom 10%
Neuronal Class	Description	Neuronal Class	Description
RIH	Ring interneuron	AVA	Command interneuron
ADL	Amphid sensory neuron	AVE	Command internueron
CEP	Head sensory neuron	RMD	Ring motor
IL2	Head sensory neuron	RME	Ring motor
OLL	Head sensory neuron	AVB	Command interneuron
AVH	Interneuron	SMD	Ring motor
URY	Head sensory neuron	AVD	Command interneuron

Table 4. Neuronal class and type for those nodes contained within the top and bottom 10% after reordering the *C. elegans* network based on the walk-based measure with $F(A) = (I - \delta A)^{-1}$ and $\delta = 0.025$.

and motor neurons. This suggests that they should lie in the middle of the global hierarchy. However, we note that the majority of postsynaptic connections made by these neurons are with motor neurons outside of the local subnetwork studied here, which justifies their placement within this subnetwork.

7. Discussion

This work considered a range of measures and algorithms relevant to discovering and quantifying hierarchical structure in a given network. The basic out-degree minus in-degree ordering (2.6) has the benefit of being the exact solution to an appropriate discrete optimization problem. Also, its connection to a random-graph model allows it to be incorporated into a likelihood test (3.5) for hierarchical versus nonhierarchical structure. Walk-based measures of the form (4.2) directly generalize this approach and have the appeal of taking a more global view of the network topology. Also, as shown in Figure 5, they can be used in conjunction with hierarchy measures to fine-tune parameters in the algorithm.

These new reordering approaches use extremely simple combinatorics, and hence they scale favorably for large, sparse networks. If the expansion (4.1) is truncated after a finite number K of terms (for example, $K = 10$), then the walk-based ranking (4.2) requires only K sparse matrix–vector multiplications. This gives an overall complexity proportional to the number of edges. The degree-based ordering is comparable with the use of a single term in (4.1) and also has complexity proportional to the number of edges.

Overall, we believe that this is a very promising methodology for discovering and quantifying hierarchy in large, complex, directed networks.

References

[Bang-Jensen and Gutin 02] Jorgen Bang-Jensen and Gregory Gutin. *Digraphs: Theory, Algorithms and Applications.* New York: Springer, 2002.

[Barnardet al. 95] S. T. Barnard, A. Pothen, and H. D. Simon. "A Spectral Algorithm for Envelope Reduction in Sparse Matrices." *Numerical Linear Algebra* 2 (1995), 317–334.

[Borgatti 05] S. P. Borgatti. "Centrality and Network Flow." *Social Networks* 27 (2005) 55–71.

[Chen et al. 07] F. Chen, Z. Chen, Z. Liu, L. Xiang, and Z. Yuan. "Finding and Evaluating the Hierarchical Structure in Complex Networks." *Journal of Physics A: Mathematical and Theoretical* 40 (2007), 5013–5023.

[Chung 97] Fan Chung. *Spectral Graph Theory.* Providence: American Mathematical Society, 1997.

[Costa et al. 07] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. "Characterization of Complex Networks: A Survey of Measurements." *Advances in Physics* 56:1 (2007), 167–242.

[Crofts and Higham 09] J. J. Crofts and D. J. Higham. "A Weighted Communicability Measure Applied to Complex Brain Networks." *Journal of the Royal Society Interface* 6:33 (2009), 411–414.

[Crofts et al. 10] J. J. Crofts, E. Estrada, D. J. Higham, and A. Taylor. "Mapping Directed Networks." *Electronic Transactions in Numerical Analysis* 37 (2010), 337–350.

[Durbin 87] Richard Michael Durbin. "Studies on the Development and organization of the Nervous System of *Caenorhabditis elegans.*" PhD thesis, University of Cambridge, 1987.

[Estrada and Hatano 07] E. Estrada and N. Hatano. "Statistical-Mechanical Approach to Subgraph Centrality in Complex Networks." *Chemical Physics Letters* 439 (2007), 247–251.

[Estrada and Hatano 08] E. Estrada and N. Hatano. "Communicability in Complex Networks." *Physical Review E* 77 (2008), 036111.

[Estrada and Higham 10] Ernesto Estrada and Desmond J. Higham. "Network Properties Revealed through Matrix Functions." *SIAM Review* 52 (2010), 696–714.

[Estrada and Rodríguez-Velázquez 05] E. Estrada and J. A. Rodríguez-Velázquez. "Subgraph Centrality in Complex Networks." *Physical Review E* 71 (2005), 056103.

[Estrada et al. 08] E. Estrada, D. J. Higham, and N. Hatano. "Communicability and Multipartite Structure in Complex Networks at Absolute Temperatures." *Physical Review E* 78 (2008), 026102.

[Estrada et al. 09] E. Estrada, D. J. Higham, and N. Hatano. "Communicability Betweenness in Complex Networks." *Physica A* 388 (2009), 764–774.

[Fortunato 10] S. Fortunato. "Community Detection in Graphs." *Physics Reports* 486 (201), 75–174.

[Girvan and Newman 02] M. Girvan and M. E. Newman. "Community Structure in Social and Biological Networks." *Proceedings of the National Academy of Sciences* 99:12 (2002), 7821–7826, 2002.

[Grindrod 02] P. Grindrod. "Range-Dependent Random Graphs and Their Application to Modeling Large Small-World Proteome Datasets." *Physical Review E* 66 (2002), 066702.

[Grindrod 03] P. Grindrod. "Modeling Proteome Networks with Range-Dependent Random Graphs." *American Journal of PharmacoGenomics* 3 (2003), 1–4.

[Grindrod 11] Peter Grindrod, Desmond J. Higham, Mark C. Parsons, and Ernesto Estrada. "Communicability across Evolving Networks." *Physical Review E* (2011), 832011, 046120.

[Higham 03] D. J. Higham. "Unravelling Small World Networks." *Journal of Computational and Applied Mathematics* 158 (2003), 61–74.

[Holme et al. 03] P. Holme, F. Liljeros, C. R. Edling, and B. J. Kim. "Network Bipartivity." *Physical Review E* 68 (2003), 056107.

[Kaiser and Hilgetag 06] M. Kaiser and C. C. Hilgetag. "Nonoptimal Component Placement, but Short Processing Paths, Due to Long-Distance Projections in Neural Systems." *PLoS Computational Biology* 2 (2006), e95.

[Katz 53] L. Katz. "A New Index Derived from Sociometric Data Analysis." *Psychometrika* 18 (1953), 39–43.

[Langville and Meyer 06] A. N. Langville and C. D. Meyer. *Google's PageRank and Beyond.* Princeton: Princeton University Press, 2006.

[Morita et al. 01] S. Morita, K. Oshio, Y. Osana, Y. Funabashi, K. Oka, and K. Kawamura. "Geometrical Stucture of the Neuronal *Caenorhabditis elegans.*" *Physica A* 298 (2001), 553–561.

[Muchnik et al. 07] L. Muchnik, R. Itzhak, S. Solomon, and Y. Louzoun. "Self-Emergence of Knowledge Trees: Extraction of Wikipedia Hierarchies." *Physical Review E* 76 (2007), 016106.

[Newman 05] M. E. J. Newman. "A Measure of Betweeness Centrality Based on Random Walks." *Social Networks* 27 (2005), 39–54.

[Newman 10] M. E. J. Newman. *Networks: An Introduction.* Oxford: Oxford University Press, 2010.

[Page et al. 98] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. "The PageRank Citation Ranking: Bringing Order to the Web." Technical report, Stanford Digital Library Technologies Project, 1998.

[Pan et al. 10] R. K. Pan, N. Chatterjee, and S. Sinha. "Mesoscopic Organization Reveals the Constraints Governing *Caenorhabditis elegans* Nervous System." *PLoS One* 5:2 (2010), e9240.

[Ravasz and Barabási 03] E. Ravasz and A. L. Barabási. "Hierarchical Organization in Complex Networks." *Physical Review E* 67 (2003), 026112.

[Sporns and Kotter 04] O. Sporns and R. Kotter. "Motifs in Brain Networks." *PLoS Biology* 2:11 (2004), 1910–1918.

[Strang 08] Gilbert Strang. *Computational Science and Engineering.* Wellesley: Wellesley–Cambridge Press, 2008.

[Trusina et al. 04] A. Trusina, S. Maslov, P. Minnhagen, and K. Sneppin. "Hierarchy Measures in Complex Networks." *Physical Review Letters* 92:17 (2004), 178702.

[Van Driessche and Roose 95] R. Van Driessche and D. Roose. "An Improved Spectral Bisection Algorithm and Its Application to Dynamic Load Balancing." *Parallel Computing* 21 (1995), 29–48.

[Varshney et al. 11] L. R. Varshney, B. L. Chen, E. Paniagua, D. H. Hall, and D. B. Chklovski. "Structural Properties of the *Caenorhabditis elegans* Neuronal Network." *PLoS Computational Biology* 7:2 (2011), e1001066.

[Zhou et al. 06] C. Zhou, L. Zemanová, G. Zamora, C. C. Hilgetag, and J. Kurths. "Hierarchical Organization Unveiled by Functional Connectivity in Complex Brain Networks." *Physical Review Letters* 97 (2006), 238103.

Jonathan J. Crofts, School of Science and Technology, Nottingham Trent University, Nottingham, NG11 8NS UK (jonathan.crofts@ntu.ac.uk)

Desmond J. Higham, Department of Mathematics and Statistics, University of Strathclyde, Glasgow, G1 1XA UK (d.j.hingham@strath.ac.uk)