Reviewer 1:

-- Ref1.0 – General comment --

Reviewer Comment	The authors present an alternative method for maximizing hierarchical relationships within directed networks, both human-made and biological. The method is elegant in design, can reproduce results of previously described automatic methods such as Vertex Sort, but provides additional information, notably probabilities of nodes existing at different levels of a hierarchic network. The method described may be of quite general use to those interested in exploring the properties of hierarchical networks, but there are a few technical issues that the authors must consider before this manuscript should be considered for publication. One in particular is crucial, this issue of generating random networks.
Author Response	We thank the reviewer for carefully reading our manuscript and provide very detailed comments. Please find our point-by-point responses below.
Excerpt From Revised Manuscript	

-- Ref1.1.1 – Method 1: HS equation --

Reviewer Comment	I think that the hierarchical score (HS) was well defined to indirectly penalize loops in hierarchical networks; however, this equation cannot be defined when Nu and Nd are equal to zero as in the authors example of figure 2 in case of a complete tree. Hence, I suggest to adapt the HS definition by defining HS equal to the suggested equation when Nu and Nd are different from zero and HS = ∞ when Nu and Nd are equal to zero. Same thing for the other 2 definitions for CHS and PHS.
Author Response	We think the reviewer may suggested to define HS=∞, when "Nu" and "Nh" (not Nu and Nd) are both zero (as the example in Figure 2). In fact, when Nu=Nh=0, our equation will lead to a hierarchical score of infinity. We have clarified this in the revision.
Excerpt From Revised Manuscript	In method → Construction of network hierarchy→ at the end of the 2nd paragraph (Page xxx) We add the following sentence: Specifically, when Nu=Nh=0, the network will have a HS of +∞.

-- Ref1.1.2 -- Method 2: Null model --

Reviewer Comment	To determine whether a network has a hierarchical structure that is unexpected by chance, the authors compared the network HS to that of 1,000 Erdos Renyi random networks having the same number of nodes and edges as the network of interest. First, as the authors stated,

	CHS but not HS is comparable for networks having different number of levels and since I do not know in advance this number for the networks to be compared, I think that CHS should be used instead of HS for this comparison. Second, I cannot see the importance of comparing the hierarchy score of a network to those of Erdos Reyni random networks, because: 1) I know in advance that Erdos Renyi graphs are random and do not have a hierarchical structure. 2) Erdos Renyi have a very different degree distribution from the network of interest and consequently such is an apples and oranges comparison. This comparison is performed to evaluate the significance of a property (i.e. hierarchy) of the network in question. So the generated networks representing the background or the null method should be similar in a certain way to the network of interest. In practice, I control for the degree distribution of the network. Hence, I suggest comparing CHS of the network of interest with CHSs of random networks generated by preserving nodes degree distribution by rewiring the edges of the original network a large number of times (e.g. 10,000).
Author Response	We thank the referee for this comments and he's right that the null is important and we've certainly thought this when developed the methods for p-value calculation. However, his/her comment presupposes that degree distribution and power are the key statistics. But it's been shown that in directed networks such as the regulatory one the statistic most correlated with phenotype is betweenness and bottleneckness (Yu et al. 2007). Hence it would be a better statistic to preserve in randomization than hubbiness. However, it would be very hard to do this. Moreover, degree preserving will distort the other important statistics. It is more unbiased to use a "flat (generic) null" such as the ER network for P-value calculation. In fact, when the degree-preserved permutation is used as the null model, some apparently hierarchical networks are (e.g. the yeast regulome) not significant. When the degree-preserving permutation strategy is used, the P-values are 0.42, 3e-10, 0.004, 0.31, respectively. The yeast regulome and human phosphorylome do not have a significant p-value, but apparently the two network have a hierarchical structure.
Excerpt From Revised Manuscript	

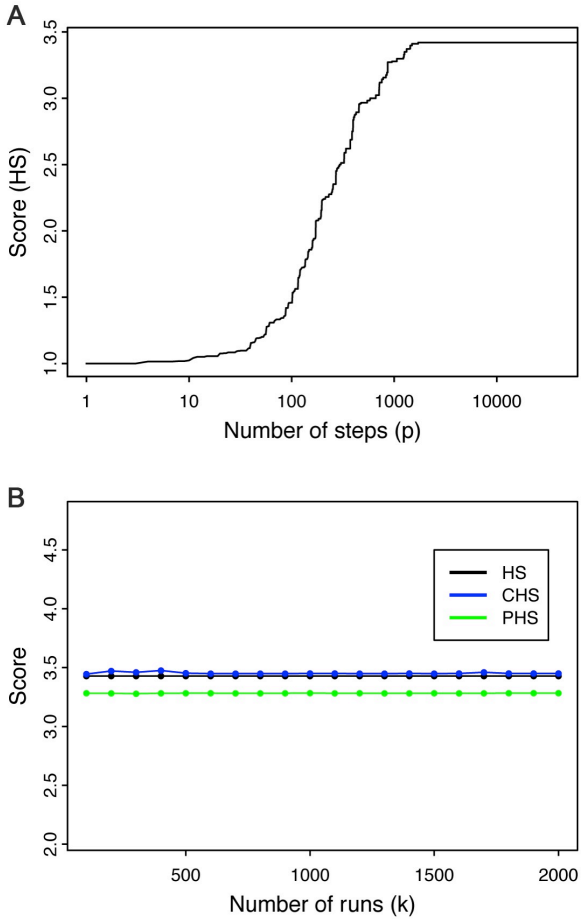
-- Ref1.1.3 -- Method 3: simulated annealing--

Reviewer Comment	The simulated iterative annealing procedure suggested in this study depends enormously on chance to maximize the hierarchy score (HS), as the node and the level in which the node should be placed are selected randomly to calculate the energy difference between the network before and after change. Obviously an exhaustive method cannot be

	used to this end because of its high cost in all terms for big networks, but I was wondering whether a semi-systematic method could be used in place of randomly choosing nodes and levels in order to ensure the largest coverage of the solutions space. For instance, making sure that all the nodes or levels were tested at least once or prioritizing the selection of certain nodes or levels.
Author Response	The reviewer made a reasonable suggestion. One way to speed up the simulated annealing procedure is to apply a more powerful cooling strategy (we are current using a simple linear cooling method that reduces temperature T gradually). Alternatively, to balance the searching space and the efficiency of optimization, we can combine HSM with other hierarchy inference method such as the VS algorithm. For example, hierarchy resulting from VS provides a good start point. Based on the hierarchy, we can use simulated annealing to further optimize the nodes with ambiguous hierarchy level assignment from VS, instead of randomly selecting nodes in each step. Such a strategy may combine the advantage of VS and HSM. In the revision, we added this idea as a new paragraph at the end of the Discussion.
Excerpt From Revised Manuscript	Discussion → last paragraph (Page xxx)

-- Ref1.1.4 -- Method 4: How to choose k and p--

Reviewer Comment	Moreover, I understand that decreasing the Temperature variable helps in accelerating the convergence of the algorithm to a global maximum, however I was wondering to what extent this is the case especially for big networks? Do the authors think that the number of repetitions of the annealing procedure should be determined as a function of the numbers of nodes and levels instead of being a fix number (1000 times) as suggested? And if they do not, why? Moreover, concerning the annealing method, it is not clear to me what does the k constant corresponds to and how it is chosen. Finally, the Simulated Annealing algorithm should be cited.
Author Response	The reviewer made a great point. It is important to determine whether the two parameters (p- the number of steps in each simulated annealing procedure, and k- the number of simulated annealing runs) are set appropriately. They can be determined by examining the p vs hs plot and k vs hs plot, respectively. The following figure shows the example for yeast regulatory network with the number of levels setting to 3 (L=3). (i) As shown in A, the hierarchy score (hs) increases gradually and reaches a fixed

	value when enough number of steps p have been performed. In this example, p=10000 will be enough to achieve the maximum value. Certainly a larger p needs to be used for a bigger network with more nodes and edges. (ii) As shown in B, for the yeast regulome when p is large enough, the probability network is fairly stable – when different k is used, ths HS, CHS and PHS are similar and similar probability P_ij (the probability of node i in level j) is obtained for all node-level pairs. We can also use plot B to determine whether a global maximum is achieved in simulated annealing. If simulated annealing results in local maximums in each run, the PHS vs. k plot will fluctuate, particularly when k is small. In summary, we can these two plots to determine p and k, and to assess whether global maximization is achieved in simulated annealing. In the revision, we add a paragraph describing how to set p and k in the method section. As suggested by the reviewer, we add the reference of the simulated annealing algorithm (Kirkpatrick et al. 1983). 
Excerpt From Revised Manuscript	Method --> Construction of network hierarchy --> last paragraph (Page xxx):

-- Ref1.2.1 – Result 1: 7-level network from VS --

Reviewer Comment	In figure 4B, the authors compared fractions of the upward, downward and horizontal edges in each of the hierarchical networks of the regulome generated by the HSM, the BFS and VS algorithms. The authors showed that the CHS score of the regulome generated by the VS is smaller than that generated by the HSM algorithm, because of the larger proportion of horizontal edges generated by the VS with 3 levels. However, it will be important to see these fractions for the VS with 7 levels, especially that theoretically the fractions of downward edges should be bigger and of the horizontal edges should be smaller than what I get in VS with L=3, which could result in a bigger CHS score. Moreover, this will clarify why the CHSs of networks generated by VS with 3 and 7 levels are comparable, especially that the latter result is not clear to me for the same reasons mentioned above.
Author Response	This is a reasonable comment. In response, we do the analysis for the VS with 7 levels. After excluding the TFs that are assigned to multiple levels, we select 78 TFs. In the yeast regulome, there are 335 (out of 580) interactions between these selected TFs. Among them, 68 (20%) are downward edges, 267 (80%) are horizontal and none is upward. If all TFs are included in the analysis, out of the 580 TF-TF interactions, 281 (42%) are downward (if the minimum level associated with regulatory node is higher than the maximum level associated with the target node) and 299 (52%) are horizontal (if the two nodes are in the same level or share a level). These results are similar to what we show in Figure 4B for VS result with level 3. We put these results in the revision.
Excerpt From Revised Manuscript	Result --> Comparison with other hierarchy construction algorithms --> 2nd paragraph (page xxx): We add a sentence at the end of the paragraph:

-- Ref1.2.2 – Result 2: p-value of networks--

Reviewer Comment	Although the authors measured the PHS for the regulome and phosphorylome, they did not estimate the significance of hierarchical structure of these 2 networks. The PHS quantifies the hierarchy underlying a given network and does not quantify the significance of this hierarchy compared to other random networks. Although the idea was presented in the methods, but no related results were reported. I think it is important to estimate the significance of hierarchical structure of the regulome and phosphorylome in order to have an idea about their divergence from random networks and the performance of the CHS score. And again, these will depend on the kind of random network chosen as I discussed above.

Author Response	We thank the reviewer for this suggestion. We calculated the significance of the four networks: yeast regulome, human regulome, yeast phosphorylome and human phosphorylome. When the ER random network is used as the null model, all of them are highly significant with P<2e-16. However, when the degree-preserving permutation strategy is used, the P-values are 0.42, 3e-10, 0.004, 0.31, respectively. The yeast regulome and human phosphorylome do not have a significant p-value, but apparently the two network harbor a hierarchical structure. As we argued in Ref1.1.2, it may not be appropriate to use the degree-preserved permutated networks as the null model. In the revision, we have added the P-value of these networks.
Excerpt From Revised Manuscript	

-- Ref1.2.3 – Result 3: regulome --

Reviewer Comment	The authors need to be more specific and refer to statistical test P-values when comparing the results generated by the HSM and VS algorithms when applied to the regulome. What do authors mean by 'tend to' while medians of boxplots are comparable, such as in the case of protein abundance, mRNA abundance and mRNA degradation? I think that such observations are overinterpreted and could hardly be supported with observing small differences in quartiles and whiskers between the compared boxplots with comparable medians. I think that biological properties of TFs in the different levels are comparable and HSM failed to show the relationship between nodes properties and the network hierarchy and this for all properties where I do not see a statistical significant difference among the 3 levels. Hence, although results in this study do not contradict what has been found by Jothi et al., they lack the statistical significance in the difference between the compared properties among the 3 levels, especially for the biological properties. I think that this should be stated more clearly when reporting the results of the regulome properties and when talking about the consistency between results generated by this and Jothi's study. Same remarks should be taken into consideration when describing results of the phosphorylome network.
Author Response	The reviewer made a reasonable suggestion. In the revision, we have rephrased many statements and deleted some sentences to avoid possible over-interpretation and confusion.
Excerpt From Revised Manuscript	Result --> Comparison with other hierarchy construction algorithms:

-- Ref1.2.4 – Result 4: Fig 6--

Reviewer Comment	The authors must support their argumentations with P-values for observed/expected values in figures 6A, 6B and 6C.
Author Response	This is a reasonable comment. We have revised Figure 6A-C. In the revised figure, we provide the p-values.
Excerpt From Revised Manuscript	

-- Ref1.2.5 – Result 5: phosphorylome --

Reviewer Comment	Finally, I'm very concerned about the analysis of the phosphorylome network. This concern is not because of anything the authors did, but rather, their sources of data are suspect. They cite two references for their data, one of which is all in vitro phosphorylation data (Ptacek, et al); always highly suspect since kinases can be quite promiscuous outside of the cell and second a protein-protein interaction screen (Breitkreutz,et al.) with kinases as baits which do not provide directed interactions. These data provide very poor information compared to those used to construct transcriptional regulatory netoworks. That's why people have shied away from phosphorylomes. I simply don't think that the biological interpretation of the constructed network can be taken seriously. I can imagine ways to construct more meaningful phosphorylomes, but it would be a lot of work and so my recommendation is that they simply remove the phosphorylome. It does not add to the story and could be highly misleading.
Author Response	We agree that the quality of the phosphorylome network may not be as good as the regulome data, but the two datasets used in this analysis were published in high-profile journals and represented the most comprehensive phosphorylation data to now. We think results from phosphorylome network can provide some biological insight with careful interpretation. As suggested by the editor, we re-interpret the results with caution but keep it in the revision. In the revision, we rephrased or deleted some sentences to avoid mis-interpretation.
Excerpt From Revised Manuscript	

-- Ref1.3.1 – Discussion 1: local vs. global --

Reviewer Comment	Authors claimed that all previous methods used to infer the hierarchical structure of a network employs a bottom-

	up approach. This is true for leaf removal and breadth first search algorithms, however this does not apply to the VS algorithm. To classify nodes in different levels, the VS algorithm applies the leaf removal algorithm on the graph and on its transpose. It then combines the 2 results to infer a global solution. Hence, I doubt that the VS algorithm could not be considered as a global method.
Author Response	We thank the reviewer for correcting this. We have revised it in the revision.
Excerpt From Revised Manuscript	Discussion → Global optimization versus local optimization → the 2nd paragraph (Page xxx):

-- Ref1.3.2 – Discussion 2: application to disconnected nets--

Reviewer Comment	When comparing the HSM to the other algorithms, I could arguably claim that other algorithms can also be applied on disconnected networks, by applying the algorithm on each disconnected proportion of the network. Hence, what is particularly useful about HSM concerning this point? Moreover, the authors claimed that results of the HSM algorithm are robust. I think that this statement should be supported by strong arguments or, more effectively, by actual results showing the robustness of HSM results on the addition of new nodes and edges. However, I agree that a major improvement of HSM beyond other algorithms is indeed the probabilistic representation of a hierarchical network.
Author Response	We thank the reviewer for the comments. We agree that the other algorithms can also be applied to disconnected network by applying them to each disconnected components. In terms of the robustness, we have performed perturbation to the yeast regulome by randomly removing 10%-50% of interactions. Our results showed 88%, 79%, 81%, 77% and 72% ~~were~~ preserve their levels in the resulting hierarchical networks (use the complete data as the reference, and set #L=3), when 10%, 20%, 30%, 40% and 50% of interactions are removed. However, since adding or deleting of nodes/edges will change the hierarchical structure of the resulting network, such a robustness analysis may not be valid. In the revision, we deleted the statement about robustness of the HSM method.
Excerpt From Revised Manuscript	

-- Ref1.3.3 – Discussion 3: GRC --

Reviewer Comment	I also agree with the authors on the outperformance of the HS over the dyadic reciprocity and Krackhardt scores in capturing the hierarchical flow, because these 2 metrics do not encapsulate information about the orientation.

	However, I think that it will be important that the authors compare the HS score to another metric similar to the HS in capturing the hierarchy flow, this metric is the global reaching coefficient (GRC) (Mones, Vicsek, & Vicsek, 2012). This comparison could be practical by comparing the HS to the GRC for the regulome and phosphorylome networks and discussing any considerable differences or concordances between the HS and the GRC for each network.
Author Response	GRC is a related metric. We thus calculated the GRC for each network and updated Table 1. As shown, the correlation between hierarchy score and GRC across these network is poor, suggesting the two metrics ~~are~~ capture different aspect of networks.
Excerpt From Revised Manuscript	

-- Ref1.3.4 – Discussion 4: kinase localization--

Reviewer Comment	Interestingly, the authors found that 5 kinases of the middle level are uniquely localized in the nucleus. Did the authors check whether these kinases differ in a certain mean from the other kinases. For instance, are they known to target a particular class of proteins?
Author Response	We have reanalyzed the cellular localization data of kinases. In the new analysis, we used the manually curated data provided by SGD database to determine the cellular localization of them. We have updated Figure 5A based on the new result. We added the cellular localization of all these 94 kinases as a new column in the Supplementary Table S2.
Excerpt From Revised Manuscript	

-- Ref1.3.5 – Discussion 5: origin of hierarchy --

Reviewer Comment	The third paragraph of the section entitled "Temporal versus spatial organization of hierarchy" simply repeats what was reported in the results section about the characteristics of the phosphorylome as a function of the network hierarchy. I suggest omitting it or discussing the results without re-reporting them. The fourth paragraph of the same section ("Temporal versus spatial organization of hierarchy") begins with "We hypothesize that this may have arisen from differences in the origin of hierarchy between the yeast regulome and phosphorylome caused by distinct evolutionary pressures that shaped them.". The pronoun "this" refers to nothing mentioned previously, hence reformulating the phrase could clarify this ambiguity.
Author Response	The reviewer's comment is reasonable. We have deleted the 3rd paragraph in the revision. To avoid over-interpretation, we also revised the description about the different origin of the hierarchical structure between regulome and phosphorylome. In the revised version, we argue that hierarchy may arise from the

	temporal and/or spatial organization of regulators. We removed the sentences that regulome hierarchy is result from "evolution" but phosphorylome hierarchy is result from spatial distribution.
Excerpt From Revised Manuscript	

-- Ref1.3.6 – Discussion 6: evolution of TFs--

Reviewer Comment	Furthermore, in this paragraph, the authors stated that "During evolution, when a new cluster of genes needed to be coordinated to response to novel selective pressure, for example, it is likely that the new regulatory TF for this gene cluster was placed in the bottom hierarchical level to avoid substantial regulatory side effects.". The authors are indirectly supposing that gene transcription is the only way to coordinate the response of a set of TFs, however I could argue that "younger" TFs could be placed at the top of the hierarchy and coordinated by another biochemical regulatory mechanism such as phosphorylation or degradation.
Author Response	The reviewer made a reasonable comment. The evolution of TF in regulating gene expression is complex and there might be various scenarios. Moreover, we agree there is no convincing evidence to show that regulome hierarchy is a consequence of evolution of TFs. In the revision, we have deleted this paragraph.
Excerpt From Revised Manuscript	

-- Ref1.3.6 – Discussion 7: GRC paper --

Reviewer Comment	References Mones, E., Vicsek, L., & Vicsek, T. (2012). Hierarchy measure for complex networks. PLoS One, 7(3), e33799. doi: 10.1371/journal.pone.0033799
Author Response	We have calculated the GRC for each network used in this paper, and added the paper into the reference.
Excerpt From Revised Manuscript	

Reviewer 2:

-- Ref2.0 – General comment --

Reviewer Comment	The submitted manuscript provides a novel method to define hierarchical layers within directed networks. This method was shown to outperform existing methods for definition of hierarchical layers. The work is of high general interest in an era of large-scale biology and opens interesting perspectives for interpretation of large-scale data in many organisms.
Author Response	We thank the reviewer for the overall positive comments on our manuscript. The following are our point-to-point responses.
Excerpt From Revised Manuscript	

-- Ref2.1 – Minor comments 1 --

Reviewer Comment	There are a few minor comments/questions: - Page 6: The authors mention that non of the nuclear kinases are in the bottom layer of the hierarchical network. In this context of data on subcellular location would the authors please specify if some proteins were annotated with dual location and if so whether or how these proteins with dual location were included in the analysis.
Author Response	The reviewer made a good suggestion. In the revision, we have update the cellular localization analysis using data from the yeast database (SGD). The database provides the cellular components associated with each protein manually curated based on literatures. Compared to the cellular localization data from large-scale experiment (Hue et al), this data is more comprehensive with high quality. We have provided information about cellular localization for all the 94 kinases as a new column in Supplementary Table S2.
Excerpt From Revised Manuscript	

-- Ref2.2 -- Minor comments 2 --

Reviewer Comment	There are a few minor comments/questions: - Page 6 same paragraph: the hierarchical phospho-network suggests "signals are transduced back to the cytoplasm by bottom level kinases". In this context it would be nice if the authors could highlight one or two examples of such a cytosolic bottom layer kinase and the inferred path to nucleus and back. Giving specific names of such proteins will help the general audience to understand and validate the results.

Author Response	According to other reviewer's comment, we have significantly revised this section to avoid possible over-interpretation. The mentioned sentences have removed in the revised version.
Excerpt From Revised Manuscript	