

Text S2: Robustness Analysis of the Mouse Gene Co-expression Networks

Jun Dong and Steve Horvath*

Dept. of Human Genetics, David Geffen School of Medicine, UCLA

Dept. of Biostatistics, School of Public Health, UCLA

*Correspondence: shorvath@mednet.ucla.edu

Abstract

This is a supplement of the article “Geometric Interpretation of Gene Co-Expression Network Analysis”. Here we illustrate our theoretical results using gene expression data from the mouse application. In particular, we study the robustness of our theoretical findings with regard to alternative methods of constructing a network. We describe the results for weighted co-expression networks constructed using different soft-thresholds $\beta \geq 1$ in

$$a_{ij} = |\text{cor}(\mathbf{x}_i, \mathbf{x}_j)|^\beta.$$

Further, we report the analogous findings for unweighted networks constructed on the basis of

$$a_{ij} = \text{Ind}(|\text{cor}(\mathbf{x}_i, \mathbf{x}_j)| \geq \tau),$$

where τ is the ‘hard’ threshold parameter, and $\text{Ind}(\cdot)$ is the indicator function taking value of 1 if the condition is true, and 0 otherwise. We provide empirical evidence that co-expression modules tend to have high eigengene factorizability and that the maximum conformity assumption is satisfied for low powers of β . Our robustness analysis shows that many of our theoretical results apply even if our underlying assumptions are not satisfied.

1 Mouse Gene Co-expression Network Application

In this supplement, we illustrate our theoretical derivations using an F2 intercross between two mouse strains C3H/HeJ and C57BL/6J. Liver gene expression data from 135 female mice were used to construct a weighted network. The biological significance of the network and its 12 modules is described in (Ghazalpour *et al.*, 2006). In this supplement and in Figure 9 of the main article, we focus on the relationships between the network concepts and a gene significance measure based on body weight. We find that many of our theoretical results hold approximately even if the expression factorizability is low and when an unweighted network is used.

We have constructed weighted networks with $\beta = 1, 2, 3, 4, 5$ and 6 , and unweighted networks with $\tau = 0.65$ and 0.5 . For the unweighted networks, we use the eigengene-based network concepts of weighted networks with $\beta = 1$ for demonstration purposes.

2 Robustness of Module Definition

Our module definition was based on the topological overlap measure (TOM) in conjunction with average linkage hierarchical clustering (refer to the Methods Supplement). In this application, we used TOM of an weighted network with power $\beta = 6$. To facilitate a comparison, we used this network module assignment for the other weighted and unweighted network analysis as well. In each of the figure below, the upper panel is the dendrogram of the average linkage hierarchical clustering method using the specific network construction parameter, and the lower panel shows genes colored by their module membership. As the figures show, our module definition is quite robust with regard to the choice of network construction methods.

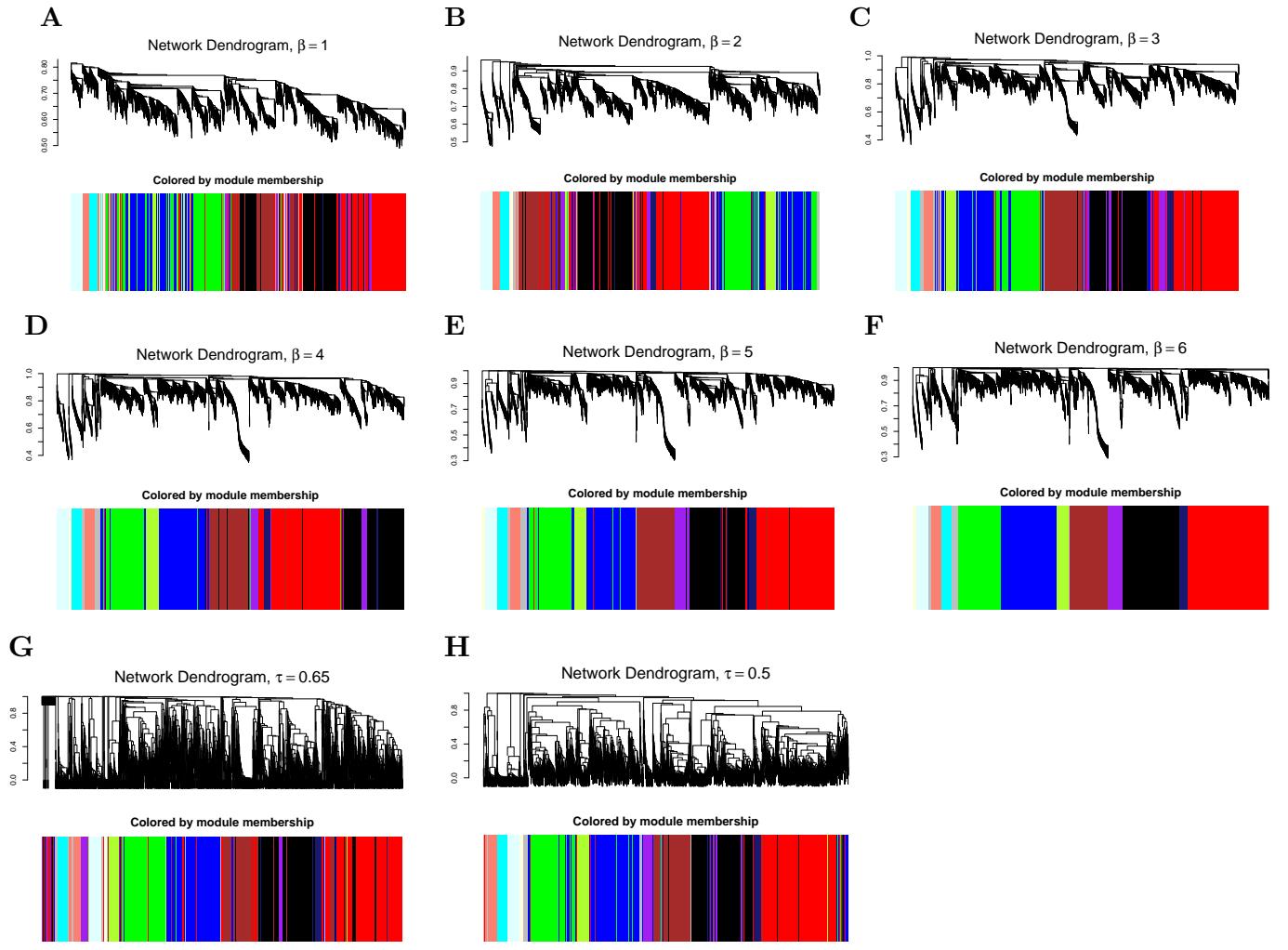


Figure 1: Robustness of module definition. In each figure, genes are colored by their module membership.

3 Summary of Robustness Analysis

To make this supplement self-contained, we repeat the following summary table from our main article. In the rest of this supplement, we provide the details on how we arrived at the R^2 values of this table.

Table 1: Robustness Analysis of the Mouse Co-expression Network. The table reports how the relation between network concepts changes as function of different soft threshold parameters β or hard thresholds used in the network construction. For each relationship and each network construction method, the table entry reports the squared correlation R^2 across the proper modules. For within module comparisons the table reports median R^2 values.

Squared Correlation R^2 Across Modules. Relation	Weighted Networks Soft Threshold β						Unweighted Net Hard Threshold τ	
	1	2	3	4	5	6	0.65	0.5
$Centralization \approx Centralization_E$	0.69	0.74	0.90	0.95	0.94	0.92	0.007	0.66
$Heterogeneity \approx Heterogeneity_E$	0.54	0.59	0.71	0.82	0.88	0.86	0.30	0.33
$ClusterCoef_i \approx ClusterCoef_E$	0.94	0.84	0.70	0.59	0.50	0.44	0.09	0.33
$ModuleSignif \approx ModuleSignif_E$	0.96	0.96	0.96	0.97	0.98	0.99	0.96	0.96
$HubGeneSignif \approx HubGeneSignif_E$	0.98	0.98	0.98	0.99	1.0	1.0	0.88	0.91
$EigengeneSignif \approx HubGeneSignif$	0.98	0.98	0.98	0.99	1.0	1.0	0.89	0.92
$ClusterCoef_i \approx (1 + (Heterogeneity)^2)^2 \times Density$	0.89	0.78	0.70	0.62	0.54	0.48	0.08	0.31
$ModuleSignif \approx \sqrt{Density} \times HubGeneSignif$	0.99	0.99	0.99	0.99	0.99	0.99	0.90	0.96
$Centralization \approx \sqrt{Density}(1 - \sqrt{Density})$	0.52	0.21	0.43	0.73	0.82	0.84	0.60	0.82
$\frac{k_{max}}{n-1} \approx \sqrt{Density}$	0.95	0.97	0.97	0.98	0.98	0.98	0.93	0.80
$K_i \approx a_{e,i}$ (median R^2)	1.0	0.99	0.98	0.96	0.95	0.94	0.74	0.86

Overall, we find that our theoretical results are highly robust in weighted networks. The relationship between the clustering coefficient and its eigengene-based analog is diminished (down to 0.44) for $\beta > 3$. The relationship between heterogeneity and its eigengene-based analog is diminished (down to 0.54 when β is low ($\beta < 3$)). The relation $Centralization^{(q)} \approx \sqrt{Density^{(q)}}(1 - \sqrt{Density^{(q)}})$ has a relatively low R^2 value (down to 0.21) for low values of $\beta \leq 3$ but the other relationships among network concepts are highly robust with respect to β . For unweighted networks, the R^2 values tend to be lower and several relationships show a marked dependency on the hard threshold τ (Table 1).

Our robustness analysis shows that many of our theoretical results apply even if our underlying assumption

tions are not satisfied. We find that the correspondence between network concepts and their eigengene-based analogs is often better in weighted networks than in unweighted networks. Further, we find that results in weighted networks tend to be more robust than those in unweighted networks with regard to changing the network construction thresholds β and τ , respectively. Thus, weighted co-expression networks are preferable over unweighted networks when a geometric interpretation of network concepts is desirable.

The correspondence between co-expression module networks and the singular value decomposition (Table 1 in the main article) can break down when a high soft threshold is used for constructing a weighted network or when dealing with an unweighted network. Thus, eigengene-based concepts do not replace network concepts when describing interaction patterns between genes.

4 Weighted Gene Co-Expression Network Results for $\beta = 1$

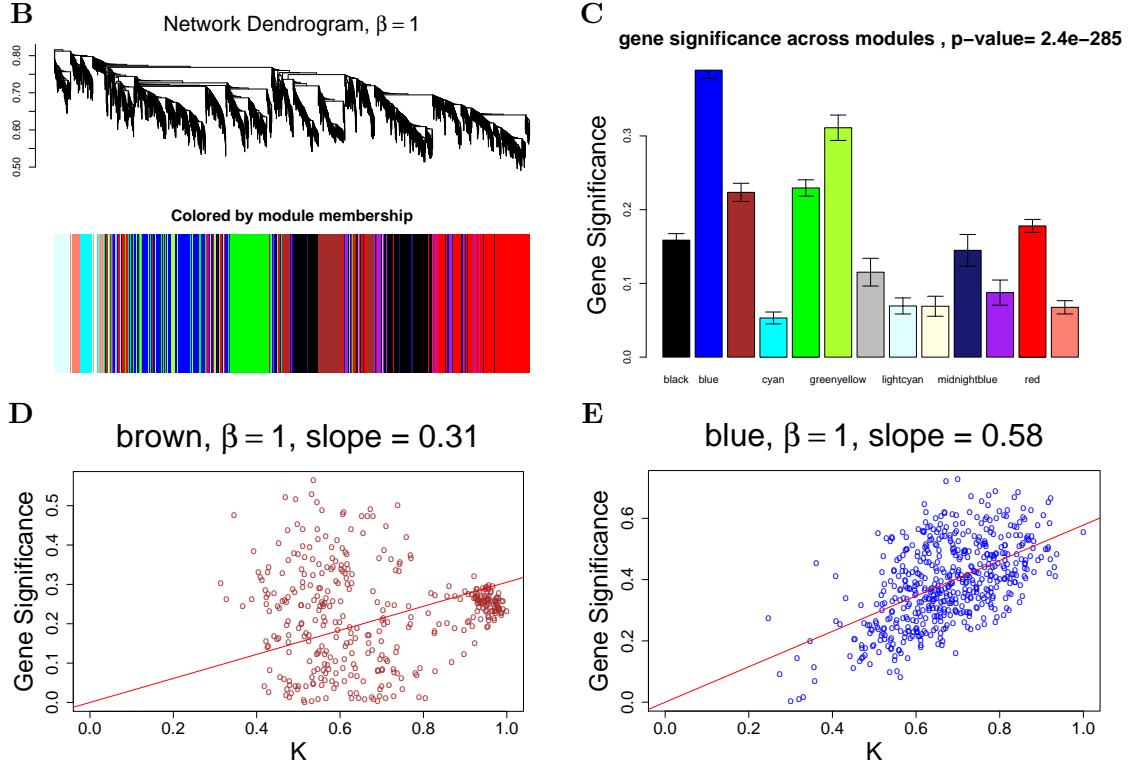


Figure 2: This figure is analogous to Figure 3 in the main article. The only difference is that we use a different dissimilarity for constructing the cluster tree in Figure B. Specifically, we use the topological overlap matrix based dissimilarity corresponding to a weighted network constructed with $\beta = 1$. Figure B depicts the hierarchical cluster tree of genes. Modules correspond to branches of the tree. The branches and module genes are assigned a color as can be seen from the color-bands underneath the tree. Grey denotes genes outside of proper modules. Figure C shows the module significance (average gene significance) of the modules. The underlying gene significance is defined with respect to the mouse body weight. Figures D and E show scatter plots of gene significance GS (y-axis) versus scaled connectivity K (x-axis) in the brown and blue module, respectively. The hub gene significance is defined as the slope of the red line, which results from a regression model without an intercept term.

Table 2: Values of network concepts for a weighted network constructed with a soft threshold of $\beta = 1$.

Module	black	blue	brown	cyan	green	greenyellow	grey	lightcyan	lightyellow	midnightblue	purple	red	salmon
Size ($n^{(q)}$)	548	534	366	96	406	121	104	119	34	84	139	772	98
Eigengene Fac. ($EF(X^{(q)})$)	0.898	0.91	0.895	0.99	0.938	0.973	0.369	0.975	0.996	0.93	0.689	0.921	0.913
$VarExplained(E^{(q)})$	0.442	0.401	0.465	0.667	0.478	0.592	0.196	0.64	0.799	0.542	0.348	0.425	0.525
$max(a_{e,i})$	0.932	0.928	0.977	0.957	0.915	0.955	0.874	0.935	0.973	0.963	0.87	0.949	0.933
Density	0.425	0.388	0.434	0.637	0.463	0.571	0.191	0.631	0.791	0.504	0.33	0.408	0.496
Density E	0.422	0.387	0.422	0.667	0.464	0.578	0.122	0.639	0.82	0.507	0.274	0.406	0.49
Centralization	0.181	0.188	0.2	0.12	0.158	0.15	0.0639	0.132	0.0781	0.17	0.11	0.197	0.147
Centralization E	0.185	0.192	0.215	0.121	0.161	0.153	0.188	0.114	0.079	0.187	0.186	0.199	0.17
Heterogeneity	0.197	0.181	0.27	0.101	0.169	0.175	0.18	0.101	0.0591	0.236	0.205	0.205	0.22
Heterogeneity E	0.218	0.193	0.323	0.101	0.174	0.18	0.786	0.103	0.0576	0.286	0.527	0.216	0.289
Mean(ClusterCoef)	0.47	0.418	0.51	0.671	0.494	0.606	0.356	0.646	0.797	0.569	0.453	0.448	0.556
ClusterCoef E	0.462	0.416	0.513	0.674	0.492	0.611	0.317	0.647	0.802	0.587	0.444	0.445	0.569
ModuleSignif	0.159	0.389	0.223	0.0531	0.229	0.311	0.115	0.0694	0.069	0.145	0.0875	0.178	0.0676
ModuleSignif E	0.0935	0.387	0.177	0.0167	0.219	0.303	0.0646	0.0388	0.0618	0.0998	0.00798	0.119	0.0399
HubGeneSignif	0.213	0.578	0.395	0.0603	0.305	0.384	0.152	0.0807	0.0751	0.167	0.0951	0.248	0.0815
HubGeneSignif E	0.134	0.578	0.266	0.0196	0.295	0.382	0.162	0.0456	0.0674	0.136	0.0133	0.177	0.0535
EigengeneSignif $= a_{e,t}^{(q)}$	0.144	0.623	0.272	0.0205	0.322	0.4	0.186	0.0488	0.0693	0.141	0.0153	0.187	0.0573

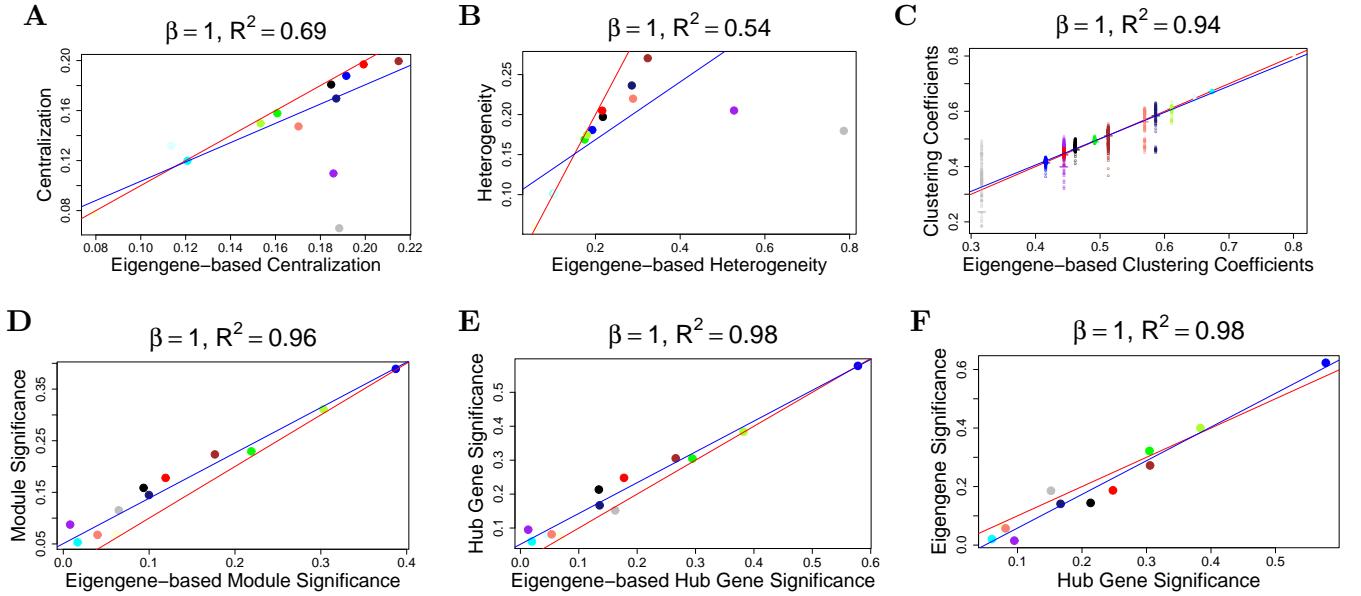


Figure 3: This figure is analogous to Figure 6 in the main article, corresponding to a weighted network constructed with a soft threshold of $\beta = 1$. It illustrates Observation 2 regarding the relationship between network concepts (y-axis) and their eigengene-based analogs (x-axis) in the mouse data. Each point corresponds to a module. Figure A: Centralization (y-axis) versus eigengene-based Centralization_E (x-axis); analogous plots for Figure B: Heterogeneity; Figure C: clustering coefficient; Figure D: module significance; and Figure E: hub gene significance. Figure F illustrates the relationship between eigengene significance and hub gene significance. The blue line is the regression line through the points representing proper modules (i.e., the grey, non-module genes are left out). While the red reference line (slope 1, intercept 0) does not always fit well, we observe high squared correlations R^2 between network concepts and their analogs. Since the grey point corresponds to the genes outside properly defined modules, we did not include it in calculations.

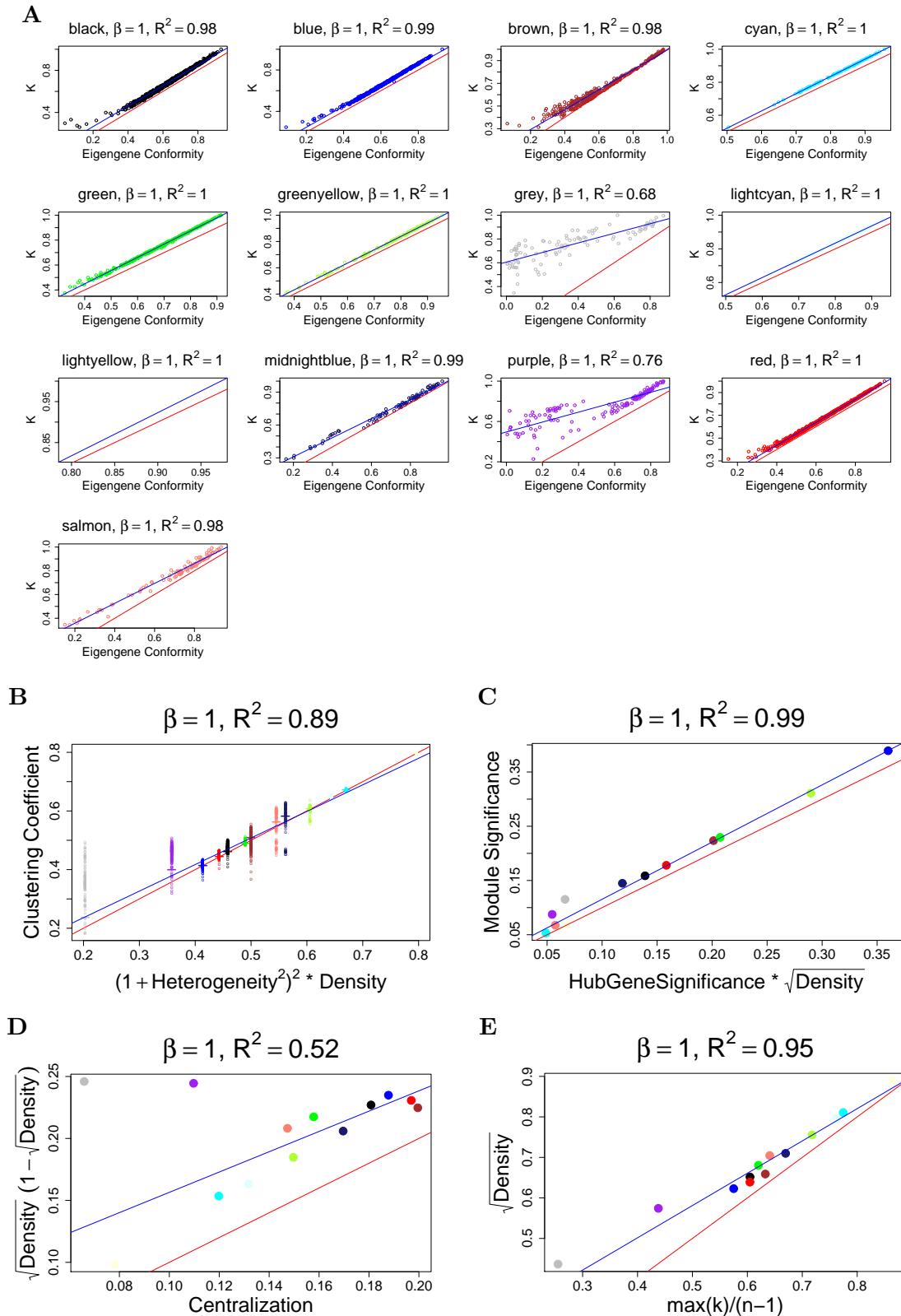


Figure 4: This figure is analogous to Figure 8 in the main article, corresponding to a weighted network constructed with a soft threshold of $\beta = 1$. It illustrates Observation 3 regarding the relationships among network concepts.

5 Weighted Gene Co-Expression Network Results for $\beta = 2$

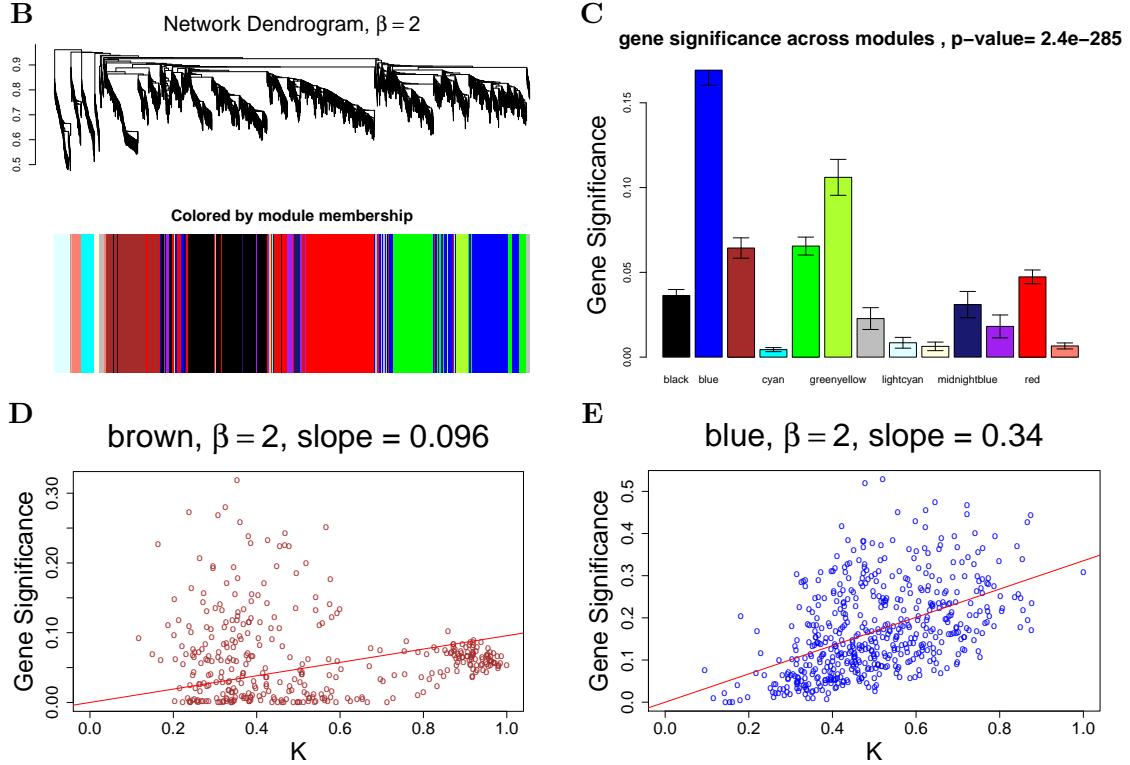


Figure 5: This figure is analogous to Figure 3 in the main article. The only difference is that we use a different dissimilarity for constructing the cluster tree in Figure B. Specifically, we use the topological overlap matrix based dissimilarity corresponding to a weighted network constructed with $\beta = 2$. Figure B depicts the hierarchical cluster tree of genes. Modules correspond to branches of the tree. The branches and module genes are assigned a color as can be seen from the color-bands underneath the tree. Grey denotes genes outside of proper modules. Figure C shows the module significance (average gene significance) of the modules. The underlying gene significance is defined with respect to the mouse body weight. Figures D and E show scatter plots of gene significance GS (y-axis) versus scaled connectivity K (x-axis) in the brown and blue module, respectively. The hub gene significance is defined as the slope of the red line, which results from a regression model without an intercept term.

Table 3: Values of network concepts for a weighted network constructed with a soft threshold of $\beta = 2$.

Module	black	blue	brown	cyan	green	greenyellow	grey	lightcyan	lightyellow	midnightblue	purple	red	salmon
Size ($n^{(q)}$)	548	534	366	96	406	121	104	119	34	84	139	772	98
Eigengene Fac. ($EF(X^{(q)})$)	0.898	0.91	0.895	0.99	0.938	0.973	0.369	0.975	0.996	0.93	0.689	0.921	0.913
VarExplained($E^{(q)}$)	0.442	0.401	0.465	0.667	0.478	0.592	0.196	0.64	0.799	0.542	0.348	0.425	0.525
$\max(a_{e,i})$	0.869	0.861	0.955	0.915	0.837	0.911	0.764	0.874	0.947	0.927	0.757	0.9	0.871
Density	0.216	0.175	0.24	0.444	0.241	0.355	0.0955	0.416	0.632	0.308	0.17	0.195	0.298
Density E	0.195	0.161	0.217	0.45	0.228	0.353	0.0387	0.413	0.658	0.298	0.122	0.181	0.279
Centralization	0.17	0.173	0.211	0.166	0.161	0.186	0.0738	0.173	0.125	0.196	0.0932	0.188	0.174
Centralization E	0.19	0.186	0.23	0.171	0.173	0.194	0.114	0.154	0.13	0.216	0.145	0.203	0.187
Heterogeneity	0.321	0.303	0.476	0.192	0.29	0.304	0.314	0.193	0.114	0.368	0.303	0.355	0.333
Heterogeneity E	0.392	0.357	0.602	0.195	0.33	0.319	1.19	0.198	0.114	0.456	0.739	0.412	0.444
Mean(ClusterCof E)	0.292	0.22	0.373	0.478	0.293	0.423	0.416	0.454	0.649	0.417	0.342	0.258	0.384
ClusterCof E	0.259	0.204	0.401	0.479	0.28	0.425	0.224	0.442	0.655	0.429	0.289	0.247	0.396
ModuleSignif	0.0363	0.169	0.0644	0.00443	0.0655	0.106	0.0227	0.00847	0.00634	0.031	0.0181	0.0473	0.0066
ModuleSignif E	0.00916	0.156	0.0344	0.000281	0.0494	0.0548	0.00676	0.00152	0.00383	0.0108	8.13e-05	0.0149	0.00173
HubGeneSignif	0.0576	0.335	0.0957	0.0054	0.105	0.15	0.0327	0.0103	0.00742	0.0341	0.0154	0.0775	0.00888
HubGeneSignif E	0.0181	0.334	0.0708	0.000385	0.0868	0.146	0.0264	0.00208	0.00454	0.0184	0.00177	0.0315	0.00286
EigengeneSignif $= a_{e,t}^{(q)}$	0.0208	0.388	0.0741	0.000421	0.104	0.16	0.0345	0.00238	0.0048	0.0199	0.00234	0.035	0.00328

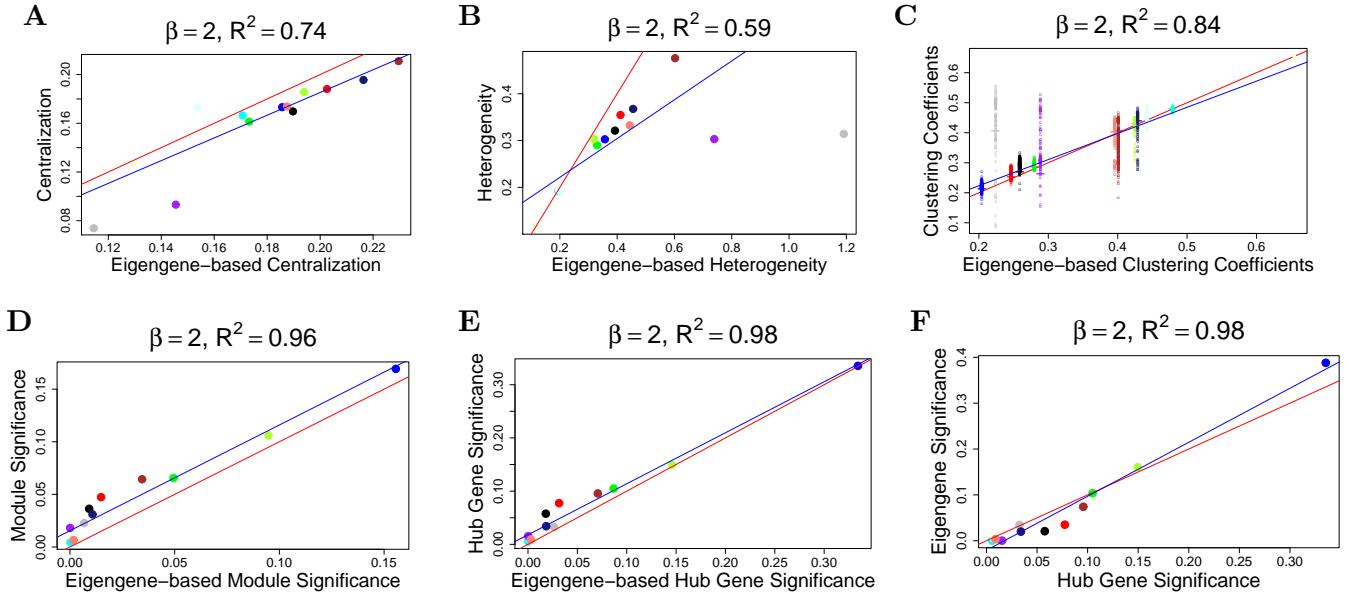


Figure 6: This figure is analogous to Figure 6 in the main article, corresponding to a weighted network constructed with a soft threshold of $\beta = 2$. It illustrates Observation 2 regarding the relationship between network concepts (y-axis) and their eigengene-based analogs (x-axis) in the mouse data. Each point corresponds to a module. Figure A: Centralization (y-axis) versus eigengene-based Centralization_E (x-axis); analogous plots for Figure B: Heterogeneity; Figure C: clustering coefficient; Figure D: module significance; and Figure E: hub gene significance. Figure F illustrates the relationship between eigengene significance and hub gene significance. The blue line is the regression line through the points representing proper modules (i.e., the grey, non-module genes are left out). While the red reference line (slope 1, intercept 0) does not always fit well, we observe high squared correlations R^2 between network concepts and their analogs. Since the grey point corresponds to the genes outside properly defined modules, we did not include it in calculations.

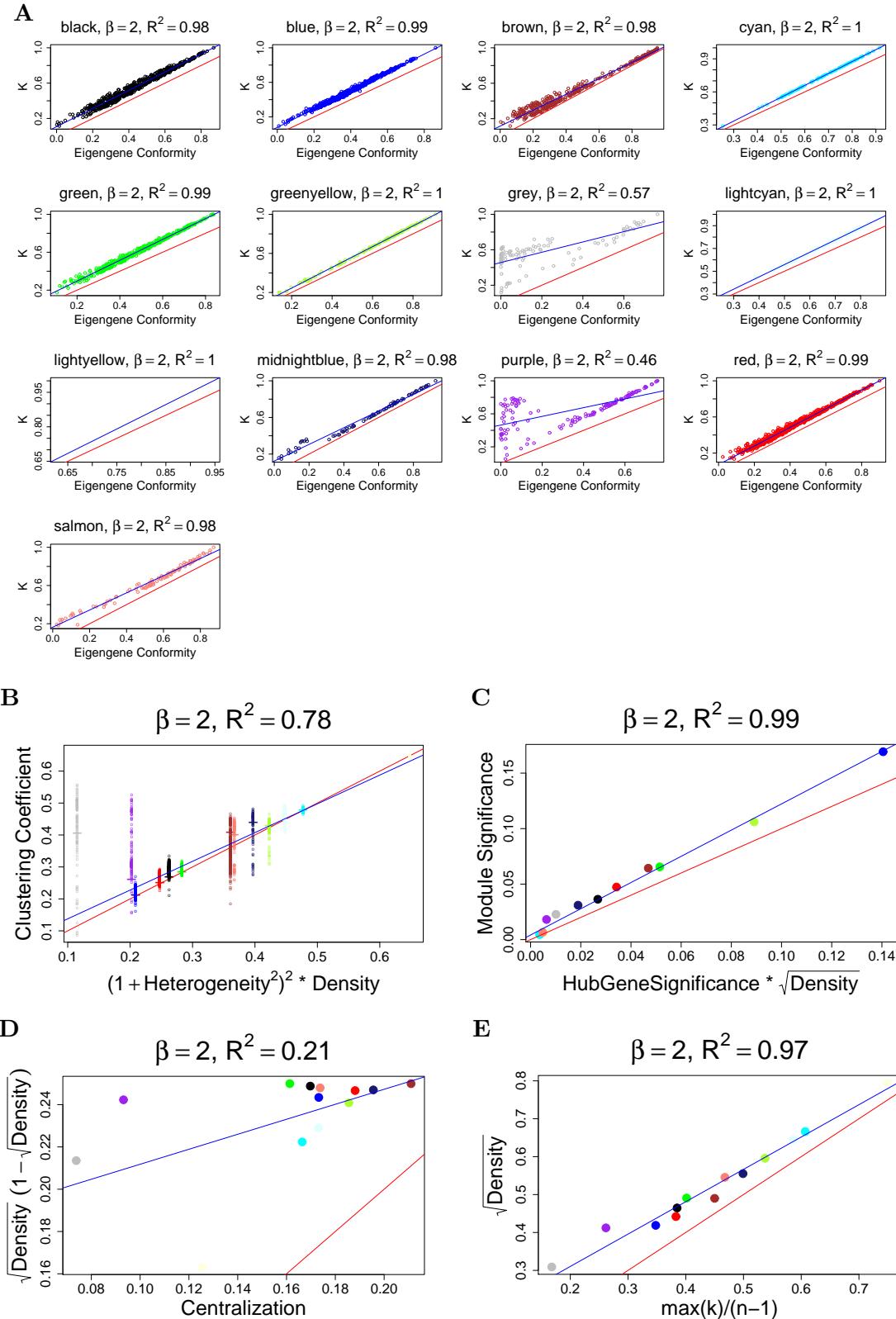


Figure 7: This figure is analogous to Figure 8 in the main article, corresponding to a weighted network constructed with a soft threshold of $\beta = 2$. It illustrates Observation 3 regarding the relationships among network concepts.

6 Weighted Gene Co-Expression Network Results for $\beta = 3$

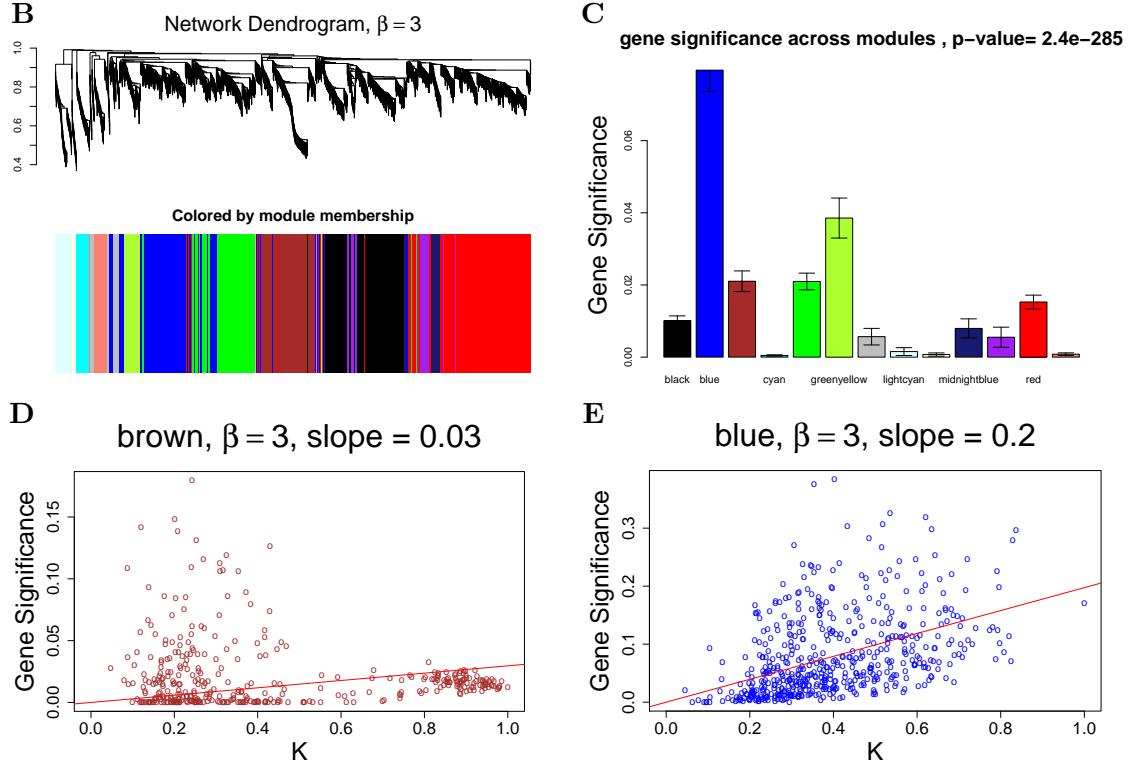


Figure 8: This figure is analogous to Figure 3 in the main article. The only difference is that we use a different dissimilarity for constructing the cluster tree in Figure B. Specifically, we use the topological overlap matrix based dissimilarity corresponding to a weighted network constructed with $\beta = 3$. Figure B depicts the hierarchical cluster tree of genes. Modules correspond to branches of the tree. The branches and module genes are assigned a color as can be seen from the color-bands underneath the tree. Grey denotes genes outside of proper modules. Figure C shows the module significance (average gene significance) of the modules. The underlying gene significance is defined with respect to the mouse body weight. Figures D and E show scatter plots of gene significance GS (y-axis) versus scaled connectivity K (x-axis) in the brown and blue module, respectively. The hub gene significance is defined as the slope of the red line, which results from a regression model without an intercept term.

Table 4: Values of network concepts for a weighted network constructed with a soft threshold of $\beta = 3$.

Module	black	blue	brown	cyan	green	greenyellow	grey	lightcyan	lightyellow	midnightblue	purple	red	salmon
Size ($n^{(q)}$)	548	534	366	96	406	121	104	119	34	84	139	772	98
Eigengene Fac. ($EF(X^{(q)})$)	0.898	0.91	0.895	0.99	0.938	0.973	0.369	0.975	0.996	0.93	0.689	0.921	0.913
$VarExplained(E^{(q)})$	0.442	0.401	0.465	0.667	0.478	0.592	0.196	0.64	0.799	0.542	0.348	0.425	0.525
$max(a_{e,i})$	0.81	0.799	0.933	0.875	0.766	0.87	0.667	0.817	0.922	0.893	0.659	0.854	0.813
Density	0.122	0.0871	0.152	0.307	0.135	0.234	0.0652	0.285	0.509	0.206	0.101	0.103	0.192
Density E	0.0964	0.071	0.13	0.309	0.118	0.226	0.017	0.272	0.531	0.19	0.0632	0.0868	0.172
Centralization	0.134	0.133	0.198	0.177	0.137	0.183	0.0712	0.175	0.152	0.19	0.0701	0.15	0.167
Centralization E	0.156	0.143	0.208	0.184	0.146	0.193	0.0718	0.159	0.16	0.206	0.105	0.165	0.171
Heterogeneity	0.422	0.405	0.666	0.275	0.392	0.406	0.4	0.281	0.167	0.466	0.402	0.484	0.422
Heterogeneity E	0.554	0.515	0.839	0.283	0.474	0.431	1.46	0.286	0.168	0.578	0.857	0.596	0.551
Mean(ClusterCof E)	0.208	0.134	0.317	0.356	0.192	0.318	0.386	0.346	0.538	0.338	0.26	0.169	0.281
ClusterCof E	0.164	0.114	0.375	0.356	0.177	0.316	0.165	0.316	0.544	0.334	0.189	0.159	0.289
ModuleSignif	0.0101	0.0795	0.021	0.000461	0.021	0.0385	0.00567	0.00154	0.000721	0.00796	0.00551	0.0153	0.00082
ModuleSignif E	0.000929	0.0644	0.00725	4.77e-06	0.0115	0.0304	0.000832	6.04e-05	0.000238	0.00121	8.97e-07	0.00193	7.75e-05
HubGeneSignif	0.0183	0.197	0.0296	0.000588	0.0384	0.0589	0.0071	0.00174	0.000908	0.00803	0.00346	0.027	0.00116
HubGeneSignif E	0.00243	0.193	0.0188	7.56e-06	0.0256	0.0558	0.00428	9.5e-05	0.000306	0.0025	2.36e-06	0.00559	0.000153
EigengeneSignif = $a_{e,t}^{(q)}$	0.00299	0.242	0.0202	8.63e-06	0.0333	0.0641	0.00641	0.000116	0.000332	0.0028	3.58e-06	0.00655	0.000188

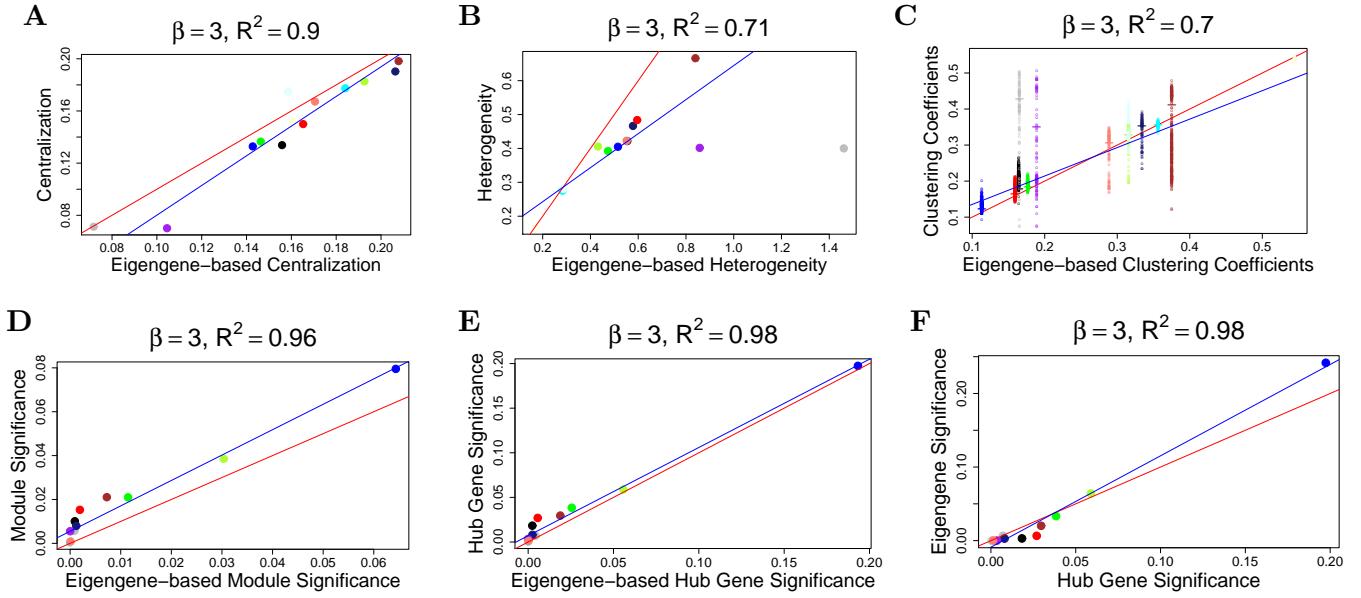


Figure 9: This figure is analogous to Figure 6 in the main article, corresponding to a weighted network constructed with a soft threshold of $\beta = 3$. It illustrates Observation 2 regarding the relationship between network concepts (y-axis) and their eigengene-based analogs (x-axis) in the mouse data. Each point corresponds to a module. Figure A: Centralization (y-axis) versus eigengene-based Centralization_E (x-axis); analogous plots for Figure B: Heterogeneity; Figure C: clustering coefficient; Figure D: module significance; and Figure E: hub gene significance. Figure F illustrates the relationship between eigengene significance and hub gene significance. The blue line is the regression line through the points representing proper modules (i.e., the grey, non-module genes are left out). While the red reference line (slope 1, intercept 0) does not always fit well, we observe high squared correlations R^2 between network concepts and their analogs. Since the grey point corresponds to the genes outside properly defined modules, we did not include it in calculations.

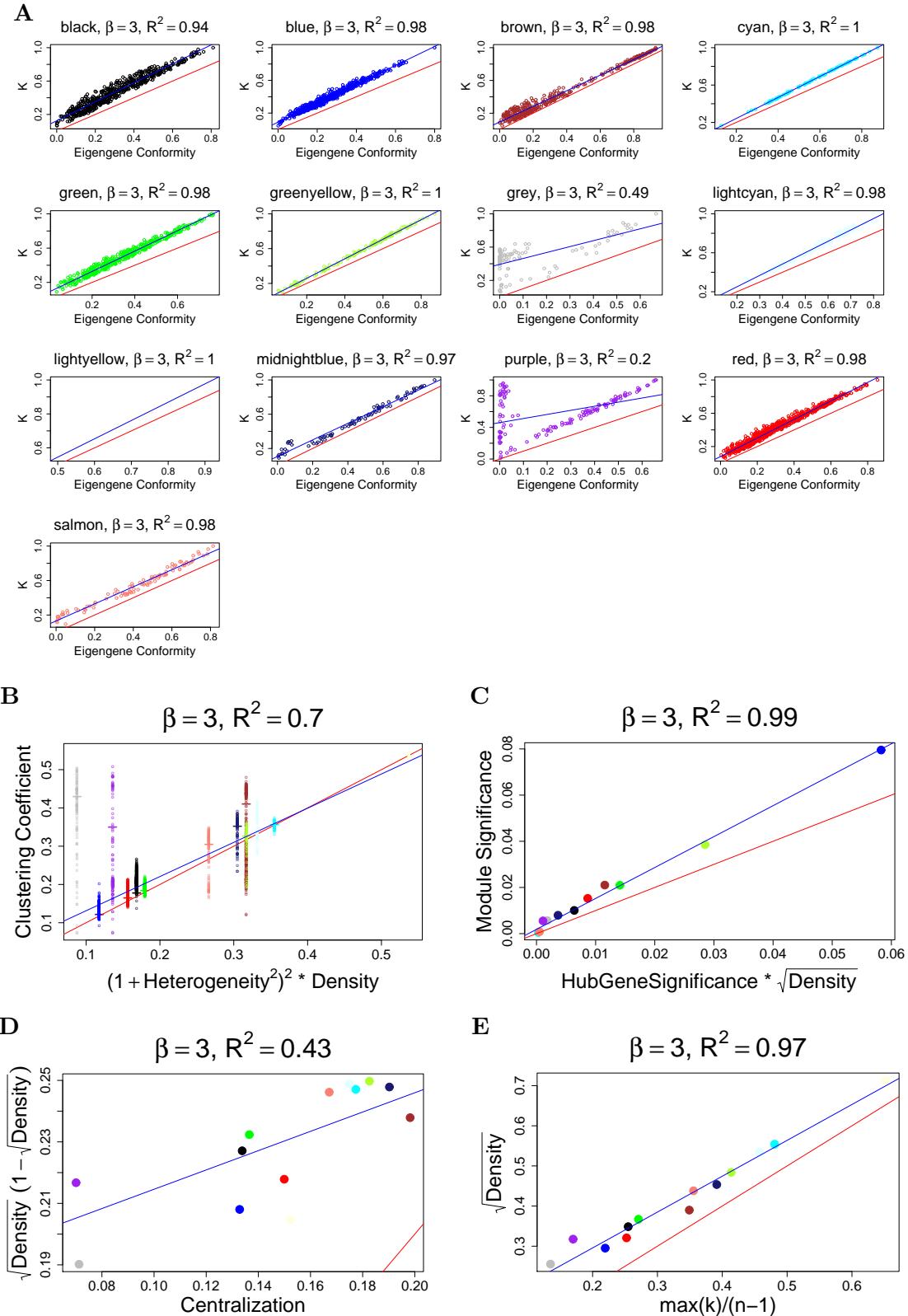


Figure 10: This figure is analogous to Figure 8 in the main article, corresponding to a weighted network constructed with a soft threshold of $\beta = 3$. It illustrates Observation 3 regarding the relationships among network concepts.

7 Weighted Gene Co-Expression Network Results for $\beta = 4$

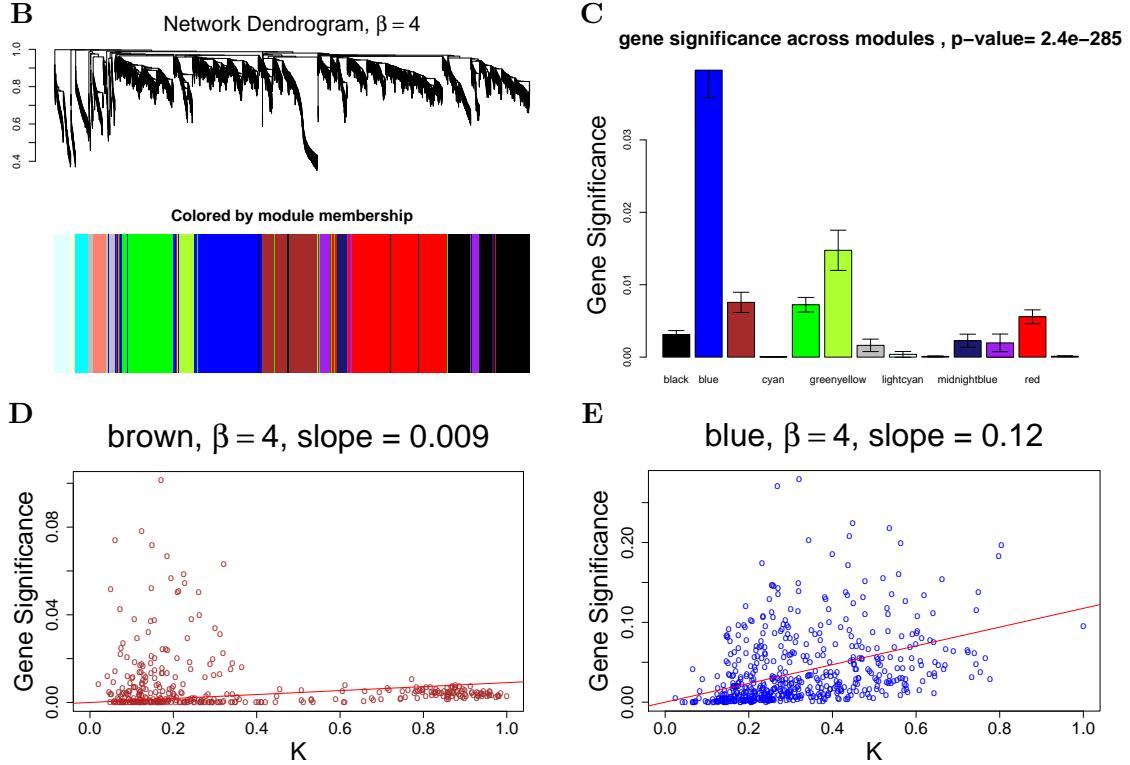


Figure 11: This figure is analogous to Figure 3 in the main article. The only difference is that we use a different dissimilarity for constructing the cluster tree in Figure B. Specifically, we use the topological overlap matrix based dissimilarity corresponding to a weighted network constructed with $\beta = 4$. Figure B depicts the hierarchical cluster tree of genes. Modules correspond to branches of the tree. The branches and module genes are assigned a color as can be seen from the color-bands underneath the tree. Grey denotes genes outside of proper modules. Figure C shows the module significance (average gene significance) of the modules. The underlying gene significance is defined with respect to the mouse body weight. Figures D and E show scatter plots of gene significance GS (y-axis) versus scaled connectivity K (x-axis) in the brown and blue module, respectively. The hub gene significance is defined as the slope of the red line, which results from a regression model without an intercept term.

Table 5: Values of network concepts for a weighted network constructed with a soft threshold of $\beta = 4$.

Module	black	blue	brown	cyan	green	greenyellow	grey	lightcyan	lightyellow	midnightblue	purple	red	salmon
Size ($n^{(q)}$)	548	534	366	96	406	121	104	119	34	84	139	772	98
Eigengene Fac. ($EF(X^{(q)})$)	0.898	0.91	0.895	0.99	0.938	0.973	0.369	0.975	0.996	0.93	0.689	0.921	0.913
VarExplained($E^{(q)}$)	0.442	0.401	0.465	0.667	0.478	0.592	0.196	0.64	0.799	0.542	0.348	0.425	0.525
$\max(a_{e,i})$	0.755	0.742	0.912	0.837	0.701	0.831	0.583	0.764	0.897	0.86	0.573	0.81	0.759
Density	0.0735	0.0464	0.106	0.217	0.0797	0.162	0.0488	0.203	0.413	0.145	0.0645	0.0581	0.129
Density E	0.0505	0.0329	0.0868	0.216	0.0664	0.15	0.00869	0.183	0.431	0.128	0.0351	0.0446	0.111
Centralization	0.103	0.0973	0.183	0.171	0.109	0.166	0.0661	0.16	0.166	0.176	0.0733	0.114	0.15
Centralization E	0.12	0.102	0.183	0.179	0.114	0.176	0.0468	0.148	0.177	0.186	0.0738	0.127	0.146
Heterogeneity	0.511	0.497	0.844	0.353	0.484	0.491	0.46	0.366	0.217	0.55	0.526	0.601	0.504
Heterogeneity E	0.707	0.672	1.03	0.366	0.61	0.526	1.65	0.369	0.221	0.68	0.945	0.772	0.641
Mean(ClusterCof E)	0.161	0.0903	0.289	0.275	0.136	0.251	0.335	0.283	0.454	0.288	0.204	0.121	0.214
ClusterCof E	0.113	0.0692	0.37	0.274	0.12	0.243	0.12	0.234	0.46	0.27	0.125	0.113	0.218
ModuleSignif	0.00312	0.0396	0.00756	5.5e-05	0.00723	0.0148	0.00163	0.000383	9.52e-05	0.00226	0.00197	0.00558	0.000122
ModuleSignif E	9.639e-05	0.0273	0.00162	8.18e-08	0.00271	0.0099	0.00011	2.41e-06	1.49e-05	0.00014	1.02e-08	0.000258	3.56e-06
HubGeneSignif	0.00662	0.118	0.00903	7.21e-05	0.0146	0.0235	0.0016	0.000369	0.000131	0.00211	0.00106	0.0101	0.000172
HubGeneSignif E	0.000326	0.112	0.00501	1.48e-07	0.00753	0.0213	0.000694	4.33e-06	2.06e-05	0.000339	3.14e-08	0.000992	8.18e-06
EigengeneSignif = $a_{e,t}^{(q)}$	0.000432	0.151	0.00549	1.77e-07	0.0107	0.0256	0.00119	5.67e-06	2.3e-05	0.000394	5.47e-08	0.00123	1.08e-05

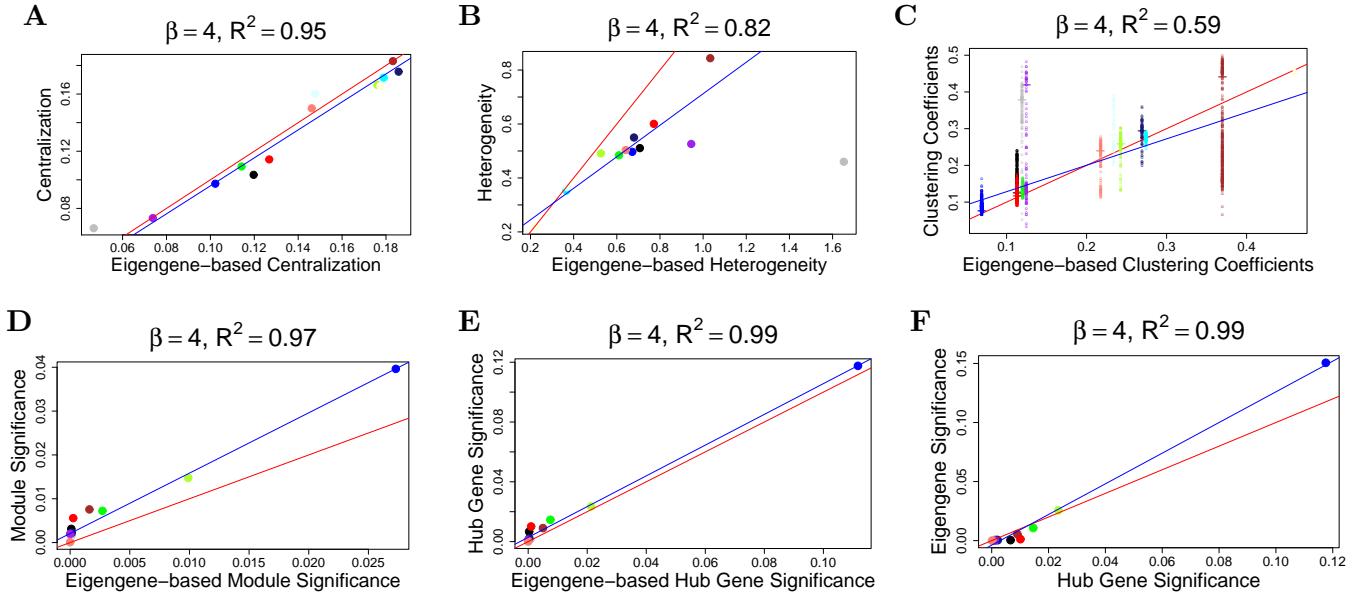


Figure 12: This figure is analogous to Figure 6 in the main article, corresponding to a weighted network constructed with a soft threshold of $\beta = 4$. It illustrates Observation 2 regarding the relationship between network concepts (y-axis) and their eigengene-based analogs (x-axis) in the mouse data. Each point corresponds to a module. Figure A: Centralization (y-axis) versus eigengene-based Centralization_E (x-axis); analogous plots for Figure B: Heterogeneity; Figure C: clustering coefficient; Figure D: module significance; and Figure E: hub gene significance. Figure F illustrates the relationship between eigengene significance and hub gene significance. The blue line is the regression line through the points representing proper modules (i.e., the grey, non-module genes are left out). While the red reference line (slope 1, intercept 0) does not always fit well, we observe high squared correlations R^2 between network concepts and their analogs. Since the grey point corresponds to the genes outside properly defined modules, we did not include it in calculations.

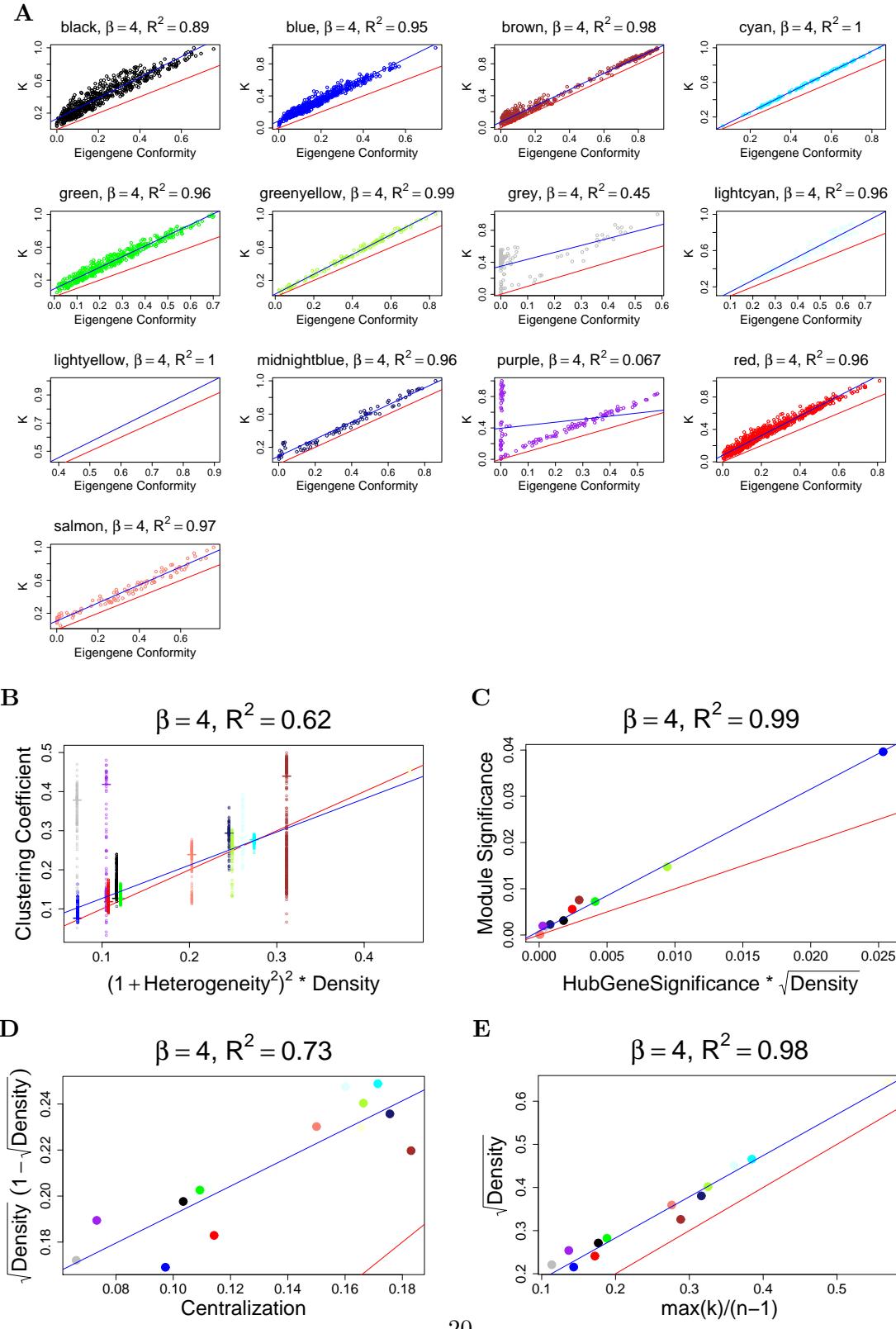


Figure 13: This figure is analogous to Figure 8 in the main article, corresponding to a weighted network constructed with a soft threshold of $\beta = 4$. It illustrates Observation 3 regarding the relationships among network concepts.

8 Weighted Gene Co-Expression Network Results for $\beta = 5$

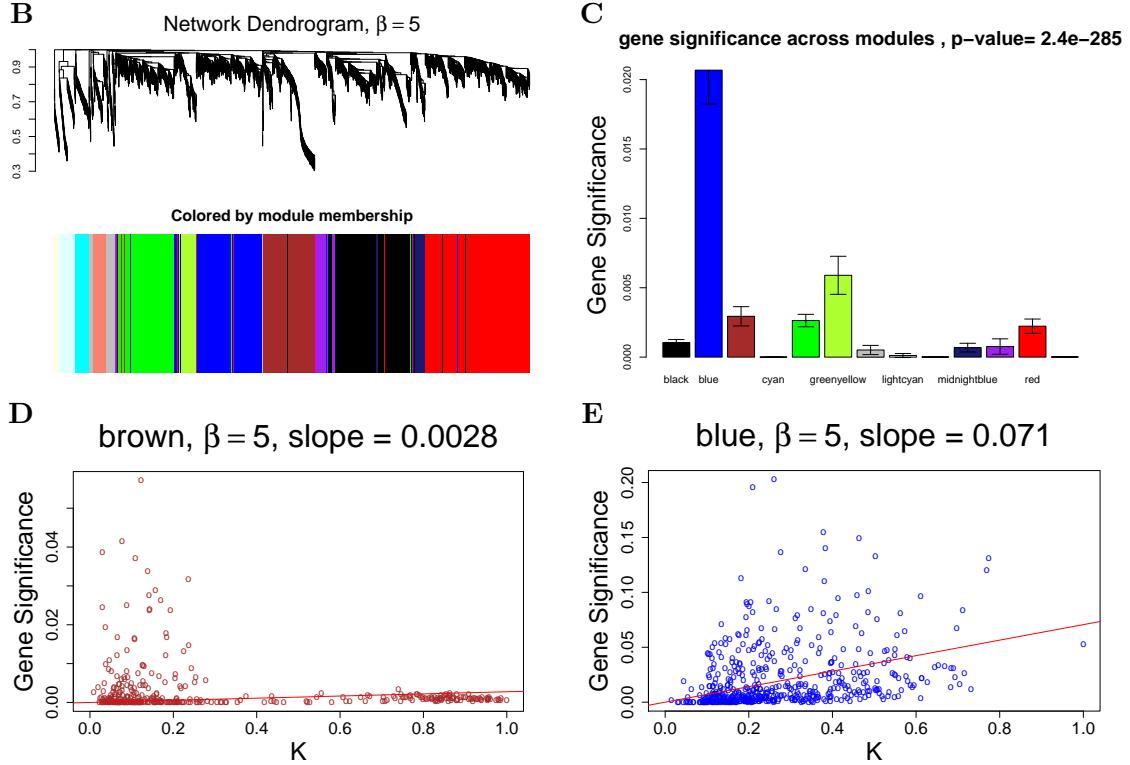


Figure 14: This figure is analogous to Figure 3 in the main article. The only difference is that we use a different dissimilarity for constructing the cluster tree in Figure B. Specifically, we use the topological overlap matrix based dissimilarity corresponding to a weighted network constructed with $\beta = 5$. Figure B depicts the hierarchical cluster tree of genes. Modules correspond to branches of the tree. The branches and module genes are assigned a color as can be seen from the color-bands underneath the tree. Grey denotes genes outside of proper modules. Figure C shows the module significance (average gene significance) of the modules. The underlying gene significance is defined with respect to the mouse body weight. Figures D and E show scatter plots of gene significance GS (y-axis) versus scaled connectivity K (x-axis) in the brown and blue module, respectively. The hub gene significance is defined as the slope of the red line, which results from a regression model without an intercept term.

Table 6: Values of network concepts for a weighted network constructed with a soft threshold of $\beta = 5$.

Module	black	blue	brown	cyan	green	greenyellow	grey	lightcyan	lightyellow	midnightblue	purple	red	salmon
Size ($n^{(q)}$)	548	534	366	96	406	121	104	119	34	84	139	772	98
Eigengene Fac. ($EF(X^{(q)})$)	0.898	0.91	0.895	0.99	0.938	0.973	0.369	0.975	0.996	0.93	0.689	0.921	0.913
VarExplained($E^{(q)}$)	0.442	0.401	0.465	0.667	0.478	0.592	0.196	0.64	0.799	0.542	0.348	0.425	0.525
$\max(a_{e,i})$	0.704	0.689	0.891	0.801	0.642	0.793	0.51	0.714	0.873	0.828	0.499	0.768	0.709
Density	0.047	0.0262	0.0801	0.156	0.0491	0.115	0.0379	0.149	0.337	0.106	0.0435	0.0346	0.0896
Density E	0.0277	0.016	0.0632	0.153	0.036	0.103	0.00482	0.125	0.352	0.0889	0.0202	0.0242	0.0732
Centralization	0.0824	0.0706	0.169	0.158	0.086	0.147	0.0605	0.14	0.17	0.159	0.0755	0.0863	0.13
Centralization E	0.0899	0.0714	0.162	0.165	0.0863	0.155	0.0313	0.131	0.185	0.163	0.0517	0.0956	0.122
Heterogeneity	0.593	0.582	1	0.426	0.568	0.565	0.508	0.451	0.264	0.624	0.677	0.708	0.58
Heterogeneity E	0.853	0.829	1.19	0.447	0.742	0.61	1.79	0.448	0.273	0.771	1.02	0.941	0.723
Mean(ClusterCof E)	0.131	0.066	0.269	0.218	0.101	0.205	0.29	0.245	0.388	0.252	0.166	0.0912	0.167
ClusterCof E	0.0826	0.0454	0.365	0.218	0.0862	0.191	0.0849	0.178	0.394	0.223	0.0841	0.0859	0.168
ModuleSignif	0.00105	0.0207	0.00294	7.15e-06	0.00264	0.0059	0.000514	0.000115	1.4e-05	0.000684	0.00076	0.00223	2.09e-05
ModuleSignif E	1.03e-05	0.0118	0.000375	1.41e-09	0.000654	0.00327	1.53e-05	9.73e-08	9.31e-07	1.65e-05	1.19e-10	3.56e-05	1.66e-07
HubGeneSignif	0.00266	0.0706	0.00276	9.58e-06	0.0057	0.00947	0.000377	9.08e-05	2.11e-05	0.000603	0.000346	0.004	2.77e-05
HubGeneSignif E	4.38e-05	0.0646	0.00133	2.91e-09	0.00222	0.00814	0.000113	1.98e-07	1.39e-06	4.6e-05	4.18e-10	0.000176	4.38e-07
EigengeneSignif = $a_{e,t}^{(q)}$	6.22e-05	0.0938	0.00149	3.63e-09	0.00345	0.0103	0.000221	2.77e-07	1.59e-06	5.56e-05	8.37e-10	0.000229	6.17e-07

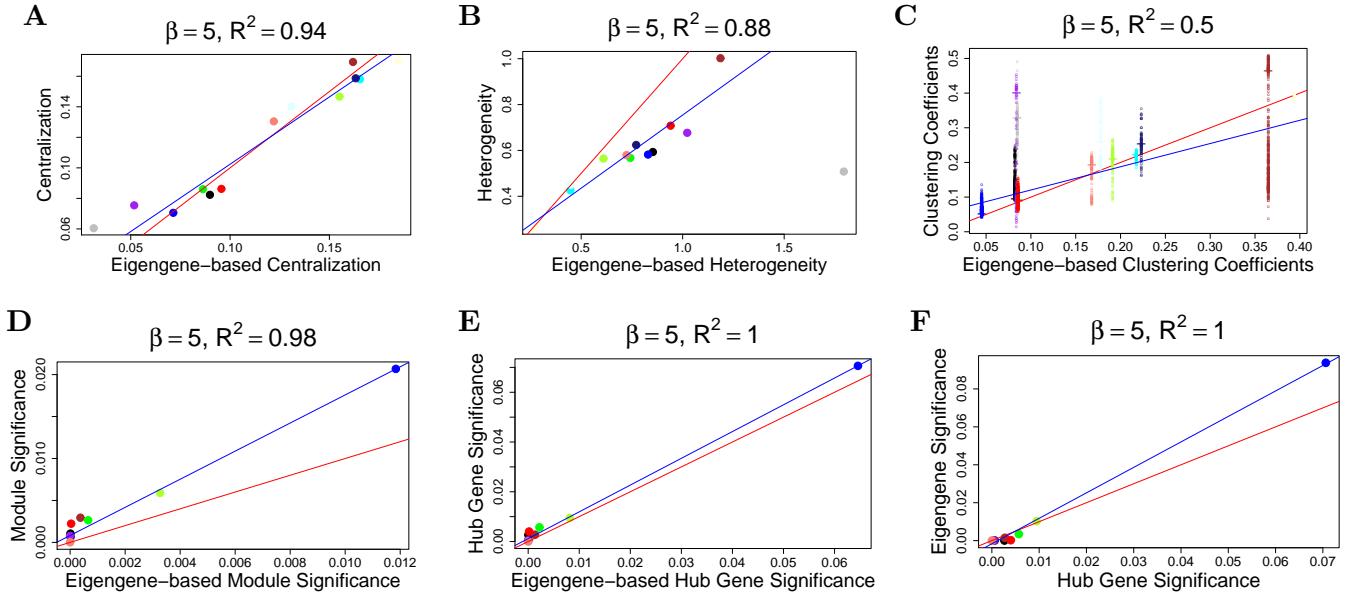


Figure 15: This figure is analogous to Figure 6 in the main article, corresponding to a weighted network constructed with a soft threshold of $\beta = 5$. It illustrates Observation 2 regarding the relationship between network concepts (y-axis) and their eigengene-based analogs (x-axis) in the mouse data. Each point corresponds to a module. Figure A: Centralization (y-axis) versus eigengene-based Centralization_E (x-axis); analogous plots for Figure B: Heterogeneity; Figure C: clustering coefficient; Figure D: module significance; and Figure E: hub gene significance. Figure F illustrates the relationship between eigengene significance and hub gene significance. The blue line is the regression line through the points representing proper modules (i.e., the grey, non-module genes are left out). While the red reference line (slope 1, intercept 0) does not always fit well, we observe high squared correlations R^2 between network concepts and their analogs. Since the grey point corresponds to the genes outside properly defined modules, we did not include it in calculations.

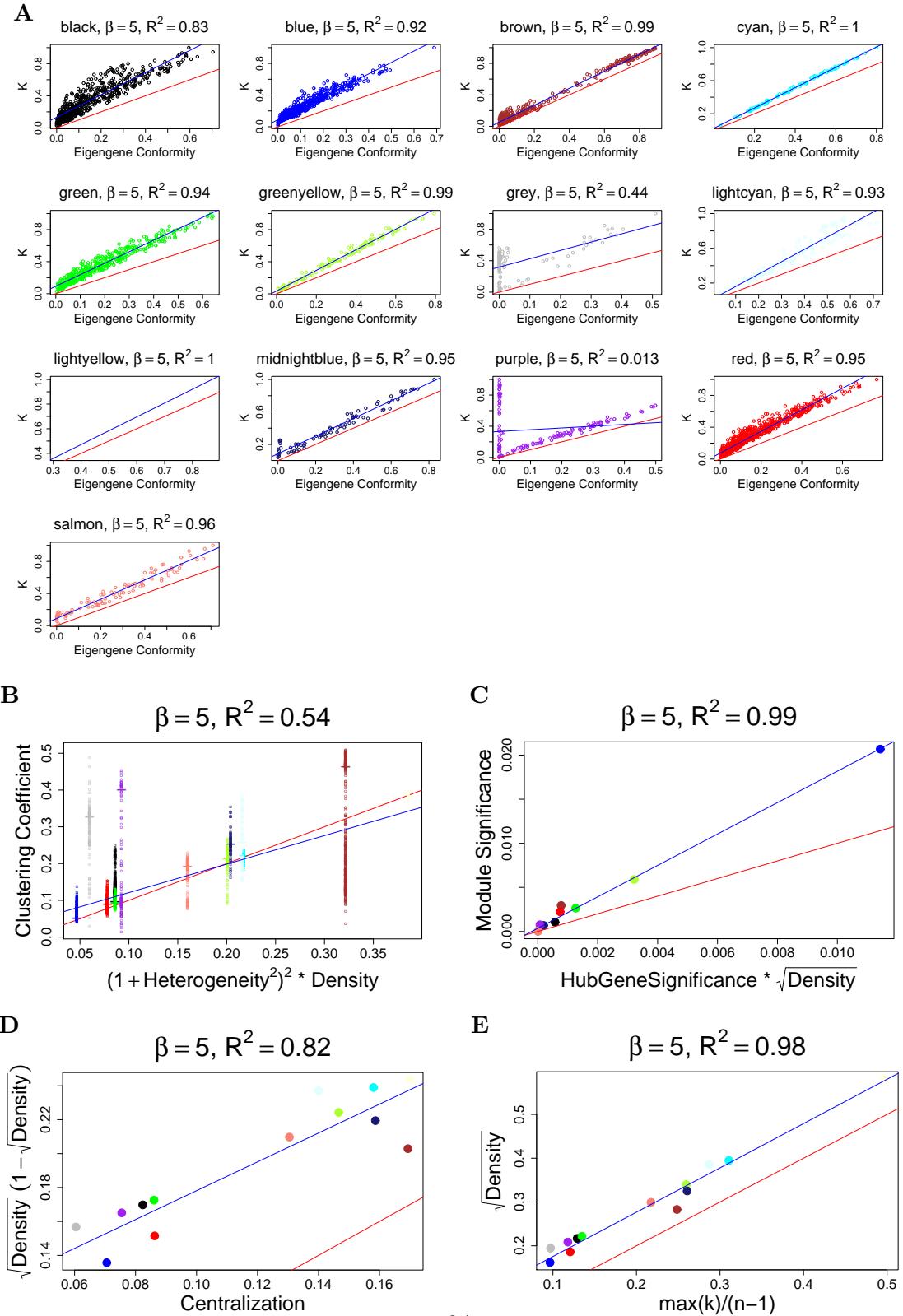


Figure 16: This figure is analogous to Figure 8 in the main article, corresponding to a weighted network constructed with a soft threshold of $\beta = 5$. It illustrates Observation 3 regarding the relationships among network concepts.

9 Weighted Gene Co-Expression Network Results for $\beta = 6$

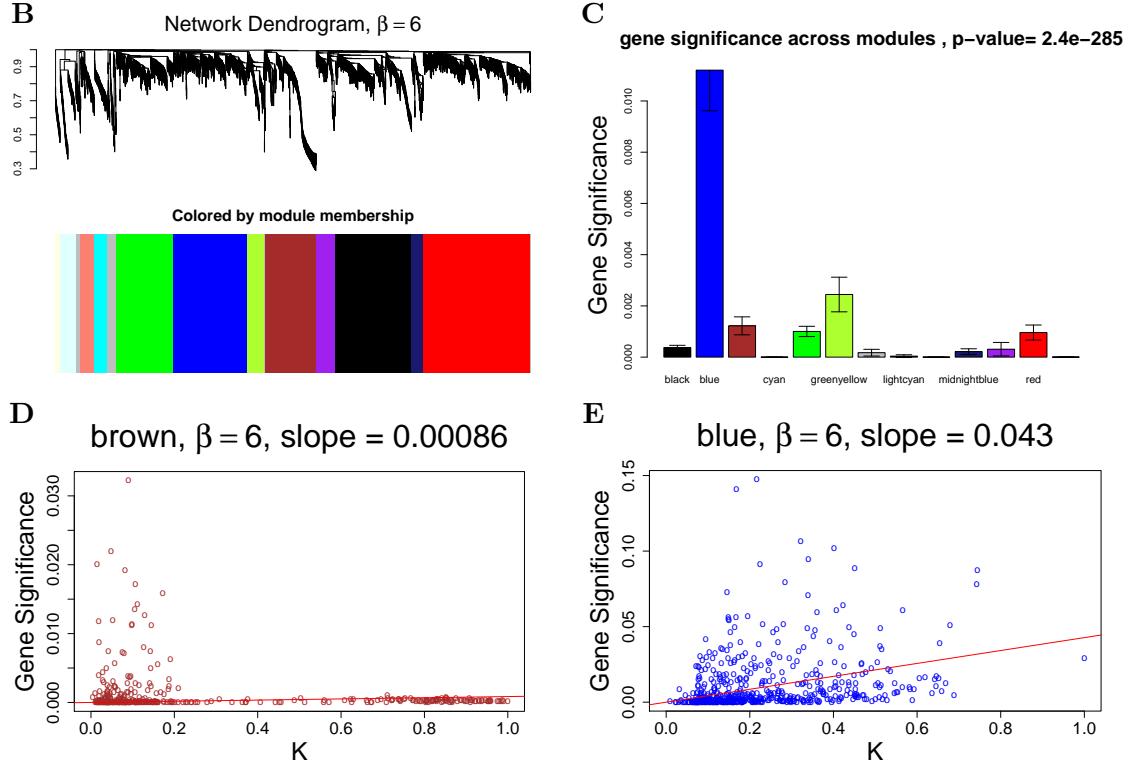


Figure 17: This figure is analogous to Figure 3 in the main article. The only difference is that we use a different dissimilarity for constructing the cluster tree in Figure B. Specifically, we use the topological overlap matrix based dissimilarity corresponding to a weighted network constructed with $\beta = 6$. Figure B depicts the hierarchical cluster tree of genes. Modules correspond to branches of the tree. The branches and module genes are assigned a color as can be seen from the color-bands underneath the tree. Grey denotes genes outside of proper modules. Figure C shows the module significance (average gene significance) of the modules. The underlying gene significance is defined with respect to the mouse body weight. Figures D and E show scatter plots of gene significance GS (y-axis) versus scaled connectivity K (x-axis) in the brown and blue module, respectively. The hub gene significance is defined as the slope of the red line, which results from a regression model without an intercept term.

Table 7: Values of network concepts for a weighted network constructed with a soft threshold of $\beta = 6$.

Module	black	blue	brown	cyan	green	greenyellow	grey	lightcyan	lightyellow	midnightblue	purple	red	salmon
Size ($n^{(q)}$)	548	534	366	96	406	121	104	119	34	84	139	772	98
Eigengene Fac. ($EF(X^{(q)})$)	0.898	0.91	0.895	0.99	0.938	0.973	0.369	0.975	0.996	0.93	0.689	0.921	0.913
VarExplained($E^{(q)}$)	0.442	0.401	0.465	0.667	0.478	0.592	0.196	0.64	0.799	0.542	0.348	0.425	0.525
$\max(a_{e,i})$	0.656	0.639	0.87	0.766	0.587	0.757	0.445	0.668	0.85	0.797	0.434	0.729	0.662
Density	0.0314	0.0155	0.0638	0.114	0.0314	0.0846	0.0301	0.113	0.277	0.0795	0.0307	0.0216	0.0636
Density E	0.0158	0.00806	0.0486	0.11	0.0269	0.0715	0.00281	0.0861	0.289	0.0635	0.0119	0.0138	0.0496
Centralization	0.0655	0.0514	0.158	0.142	0.0675	0.127	0.0548	0.121	0.168	0.142	0.0758	0.0653	0.112
Centralization E	0.067	0.0496	0.144	0.149	0.0644	0.134	0.0213	0.113	0.186	0.142	0.0362	0.0721	0.101
Heterogeneity	0.673	0.662	1.14	0.495	0.645	0.63	0.549	0.535	0.31	0.691	0.847	0.808	0.652
Heterogeneity E	0.995	0.989	1.31	0.524	0.871	0.686	1.91	0.523	0.323	0.857	1.1	1.1	0.802
Mean(ClusterCof E)	0.111	0.0512	0.252	0.177	0.0782	0.172	0.253	0.221	0.336	0.226	0.138	0.0718	0.134
ClusterCof E	0.0626	0.0315	0.355	0.177	0.0645	0.153	0.0598	0.138	0.342	0.189	0.0576	0.0679	0.133
ModuleSignif	0.000371	0.0112	0.00122	9.87e-07	0.001	0.00244	0.000173	3.86e-05	2.22e-06	0.000215	0.000369	0.00096	3.89e-06
ModuleSignif E	1.13e-06	0.00524	8.96e-05	2.46e-11	0.000161	0.00109	2.17e-06	3.95e-09	5.84e-08	1.96e-06	1.39e-12	5.04e-06	7.84e-09
HubGeneSignif	0.00112	0.0427	0.000855	1.35e-06	0.00228	0.00387	9.14e-05	2.42e-05	3.7e-06	0.000183	0.000111	0.00167	4.72e-06
HubGeneSignif E	5.88e-06	0.0373	0.000354	5.71e-11	0.000653	0.00311	1.83e-05	9.02e-09	9.38e-08	6.24e-06	5.56e-12	3.12e-05	2.34e-08
EigengeneSignif = $a_{e,t}^{(q)}$	8.97e-06	0.0584	0.000407	7.45e-11	0.00111	0.00411	4.11e-05	1.35e-08	1.1e-07	7.83e-06	1.28e-11	4.29e-05	3.54e-08

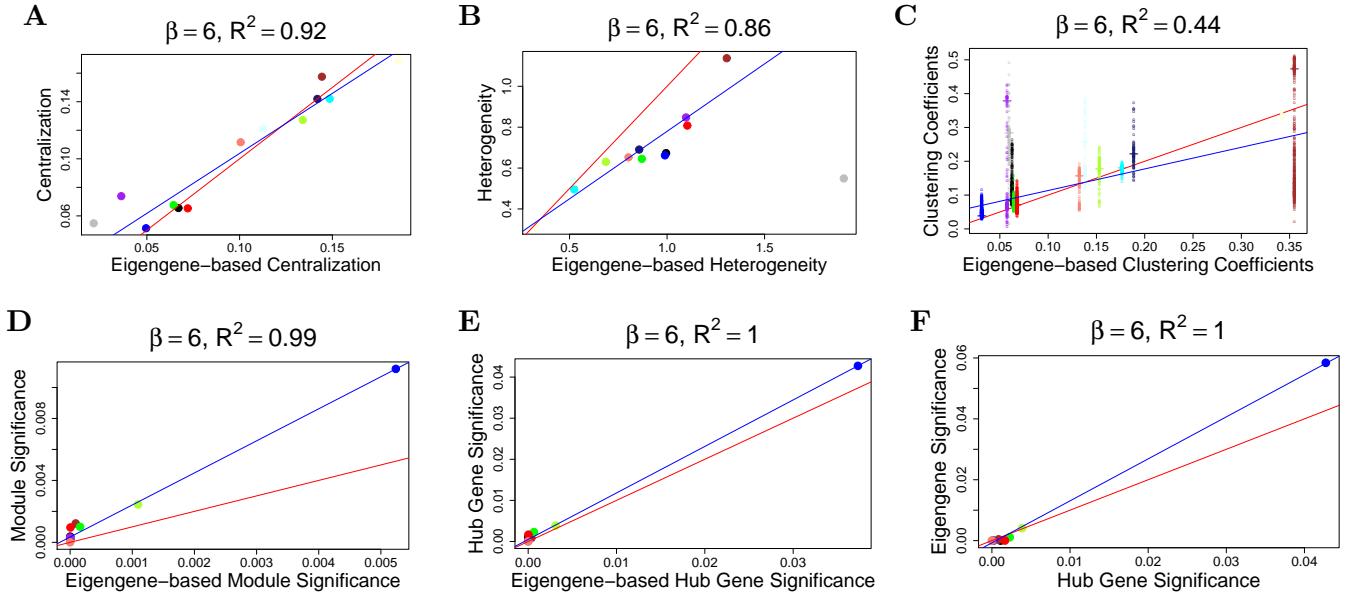


Figure 18: This figure is analogous to Figure 6 in the main article, corresponding to a weighted network constructed with a soft threshold of $\beta = 6$. It illustrates Observation 2 regarding the relationship between network concepts (y-axis) and their eigengene-based analogs (x-axis) in the mouse data. Each point corresponds to a module. Figure A: Centralization (y-axis) versus eigengene-based Centralization_E (x-axis); analogous plots for Figure B: Heterogeneity; Figure C: clustering coefficient; Figure D: module significance; and Figure E: hub gene significance. Figure F illustrates the relationship between eigengene significance and hub gene significance. The blue line is the regression line through the points representing proper modules (i.e., the grey, non-module genes are left out). While the red reference line (slope 1, intercept 0) does not always fit well, we observe high squared correlations R^2 between network concepts and their analogs. Since the grey point corresponds to the genes outside properly defined modules, we did not include it in calculations.

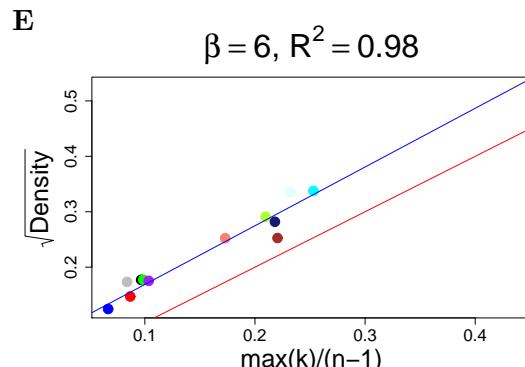
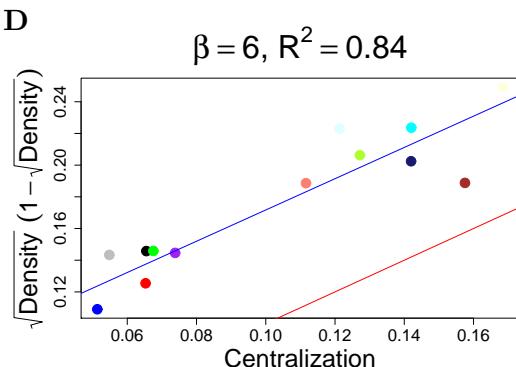
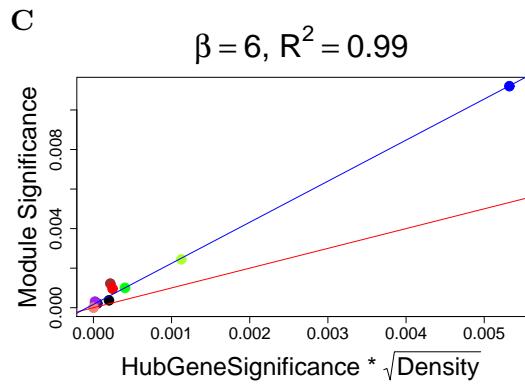
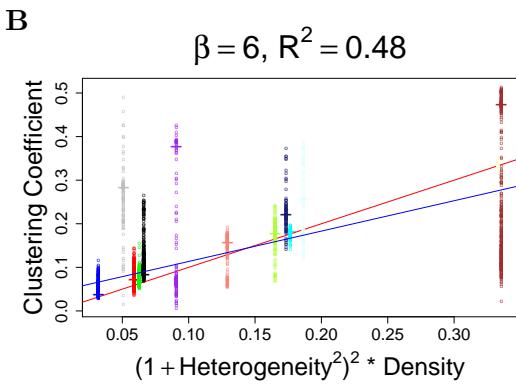
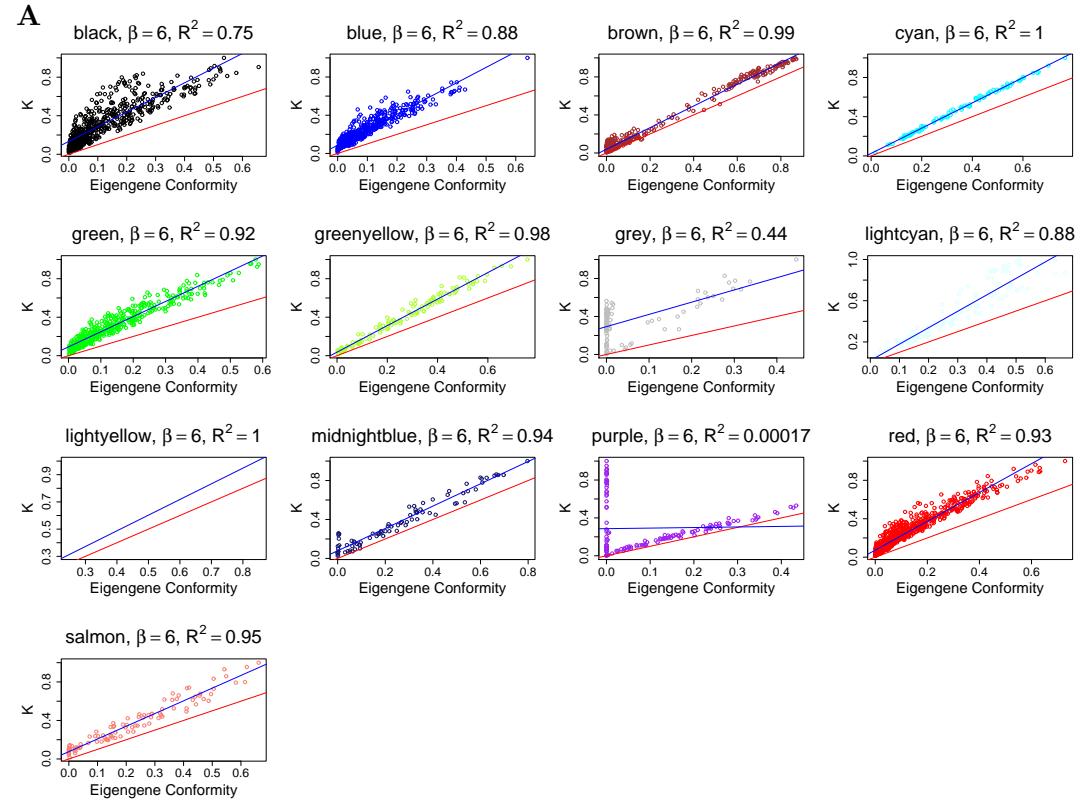


Figure 19: This figure is analogous to Figure 8 in the main article, corresponding to a weighted network constructed with a soft threshold of $\beta = 6$. It illustrates Observation 3 regarding the relationships among network concepts.

10 Unweighted Gene Co-Expression Network Results for $\tau = 0.65$

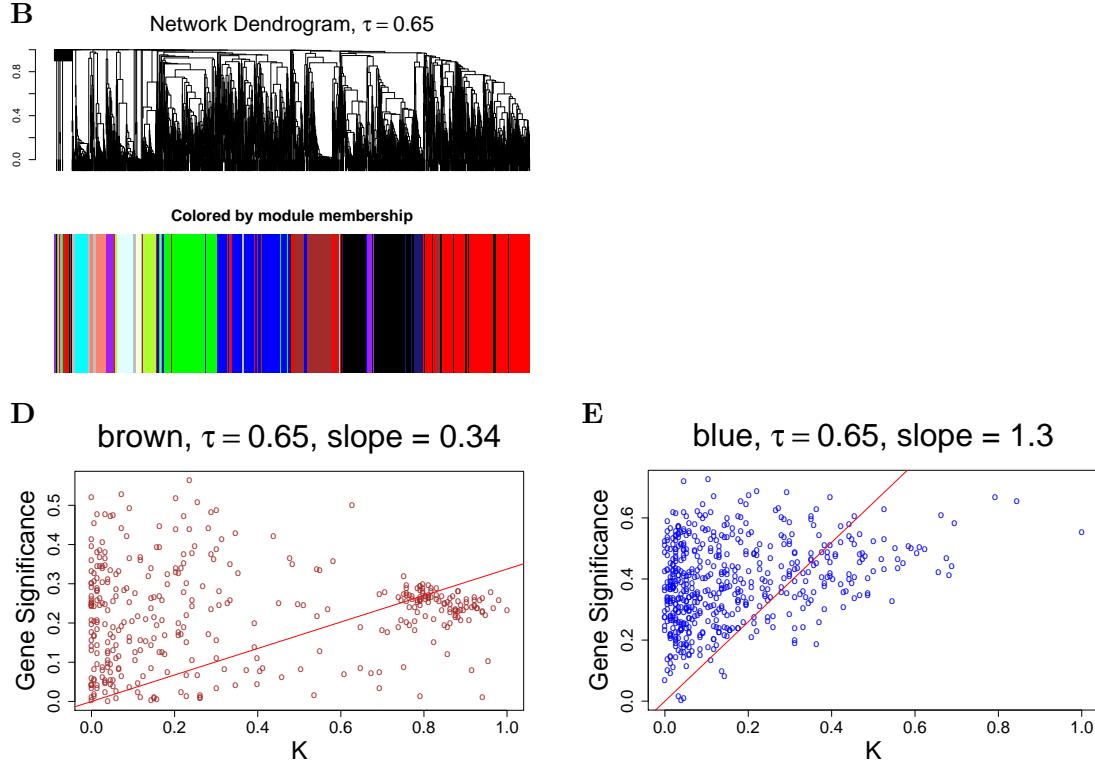


Figure 20: This figure is analogous to Figure 3 in the main article. The only difference is that we use a different dissimilarity for constructing the cluster tree in Figure B. Specifically, we use the topological overlap matrix based dissimilarity corresponding to an unweighted network constructed with $\tau = 0.65$. Figure B depicts the hierarchical cluster tree of genes. Modules correspond to branches of the tree. The branches and module genes are assigned a color as can be seen from the color-bands underneath the tree. Grey denotes genes outside of proper modules. Figure C would be identical to the case of $\beta = 1$, and thus is omitted here. Figures D and E show scatter plots of gene significance GS (y-axis) versus scaled connectivity K (x-axis) in the brown and blue module, respectively. The underlying gene significance is defined with respect to the mouse body weight. The hub gene significance is defined as the slope of the red line, which results from a regression model without an intercept term.

Table 8: Values of network concepts for an unweighted network constructed with a hard threshold of $\tau = 0.65$.

Module	black	blue	brown	cyan	green	greenyellow	grey	lightcyan	lightyellow	midnightblue	purple	red	salmon
Size ($n^{(q)}$)	548	534	366	96	406	121	104	119	34	84	139	772	98
Eigengene Fac. ($EF(X^{(q)})$)	0.898	0.91	0.895	0.99	0.938	0.973	0.369	0.975	0.996	0.93	0.689	0.921	0.913
VarExplained($E^{(q)}$)	0.442	0.401	0.465	0.667	0.478	0.592	0.196	0.64	0.799	0.542	0.348	0.425	0.525
$\max(a_{e,i})$	0.932	0.928	0.977	0.957	0.915	0.955	0.874	0.935	0.973	0.963	0.87	0.949	0.933
Density	0.125	0.0468	0.151	0.544	0.126	0.365	0.106	0.431	0.955	0.317	0.102	0.078	0.299
Density E	0.422	0.387	0.422	0.667	0.464	0.578	0.122	0.639	0.82	0.507	0.274	0.406	0.49
Centralization	0.28	0.243	0.269	0.401	0.34	0.375	0.0996	0.475	0.0473	0.341	0.234	0.3	0.389
Centralization E	0.185	0.192	0.215	0.121	0.161	0.153	0.188	0.114	0.079	0.187	0.186	0.199	0.17
Heterogeneity	0.796	1.02	0.955	0.456	0.868	0.62	0.538	0.445	0.0689	0.666	0.873	1.12	0.66
Heterogeneity E	0.218	0.193	0.323	0.101	0.174	0.18	0.786	0.103	0.0576	0.286	0.527	0.216	0.289
Mean(ClusterCof E)	0.648	0.531	0.675	0.815	0.597	0.757	0.787	0.784	0.964	0.769	0.553	0.56	0.702
ClusterCof E	0.462	0.416	0.513	0.674	0.492	0.611	0.317	0.647	0.802	0.587	0.444	0.445	0.569
ModuleSignif	0.159	0.389	0.223	0.0531	0.229	0.311	0.115	0.0694	0.069	0.145	0.0875	0.178	0.0676
ModuleSignif E	0.0935	0.387	0.177	0.0167	0.219	0.303	0.0646	0.0388	0.0618	0.0998	0.00798	0.119	0.0399
HubGeneSignif	0.306	1.3	0.337	0.0671	0.528	0.464	0.153	0.111	0.072	0.155	0.0972	0.348	0.106
HubGeneSignif E	0.134	0.578	0.266	0.0196	0.295	0.382	0.162	0.0456	0.0674	0.136	0.0133	0.177	0.0535
EigengeneSignif $= a_{e,t}^{(q)}$	0.144	0.623	0.272	0.0205	0.322	0.4	0.186	0.0488	0.0693	0.141	0.0153	0.187	0.0573

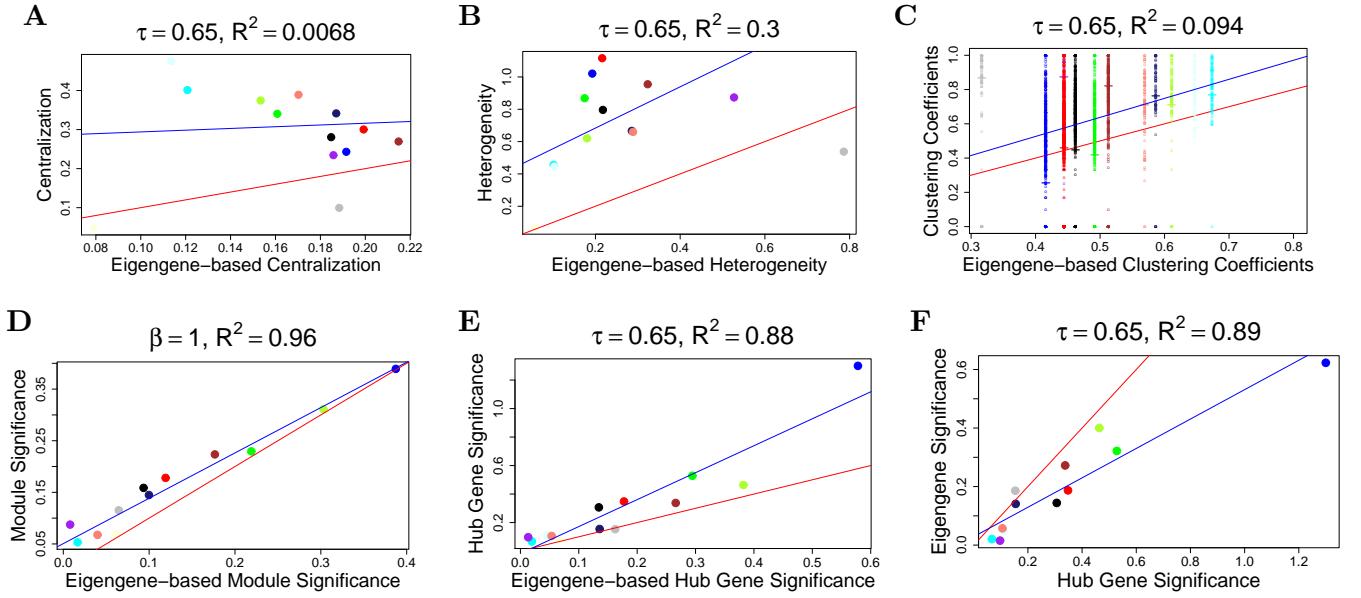


Figure 21: This figure is analogous to Figure 6 in the main article, corresponding to an unweighted network constructed with a hard threshold of $\tau = 0.65$. It illustrates Observation 2 regarding the relationship between network concepts (y-axis) and their eigengene-based analogs (x-axis) in the mouse data. Each point corresponds to a module. Figure A: Centralization (y-axis) versus eigengene-based Centralization_E (x-axis); analogous plots for Figure B: Heterogeneity; Figure C: clustering coefficient; Figure D: module significance; and Figure E: hub gene significance. Figure D is identical to the case of $\beta = 1$. Figure F illustrates the relationship between eigengene significance and hub gene significance. The blue line is the regression line through the points representing proper modules (i.e., the grey, non-module genes are left out). While the red reference line (slope 1, intercept 0) does not always fit well, we observe high squared correlations R^2 between network concepts and their analogs. Since the grey point corresponds to the genes outside properly defined modules, we did not include it in calculations.

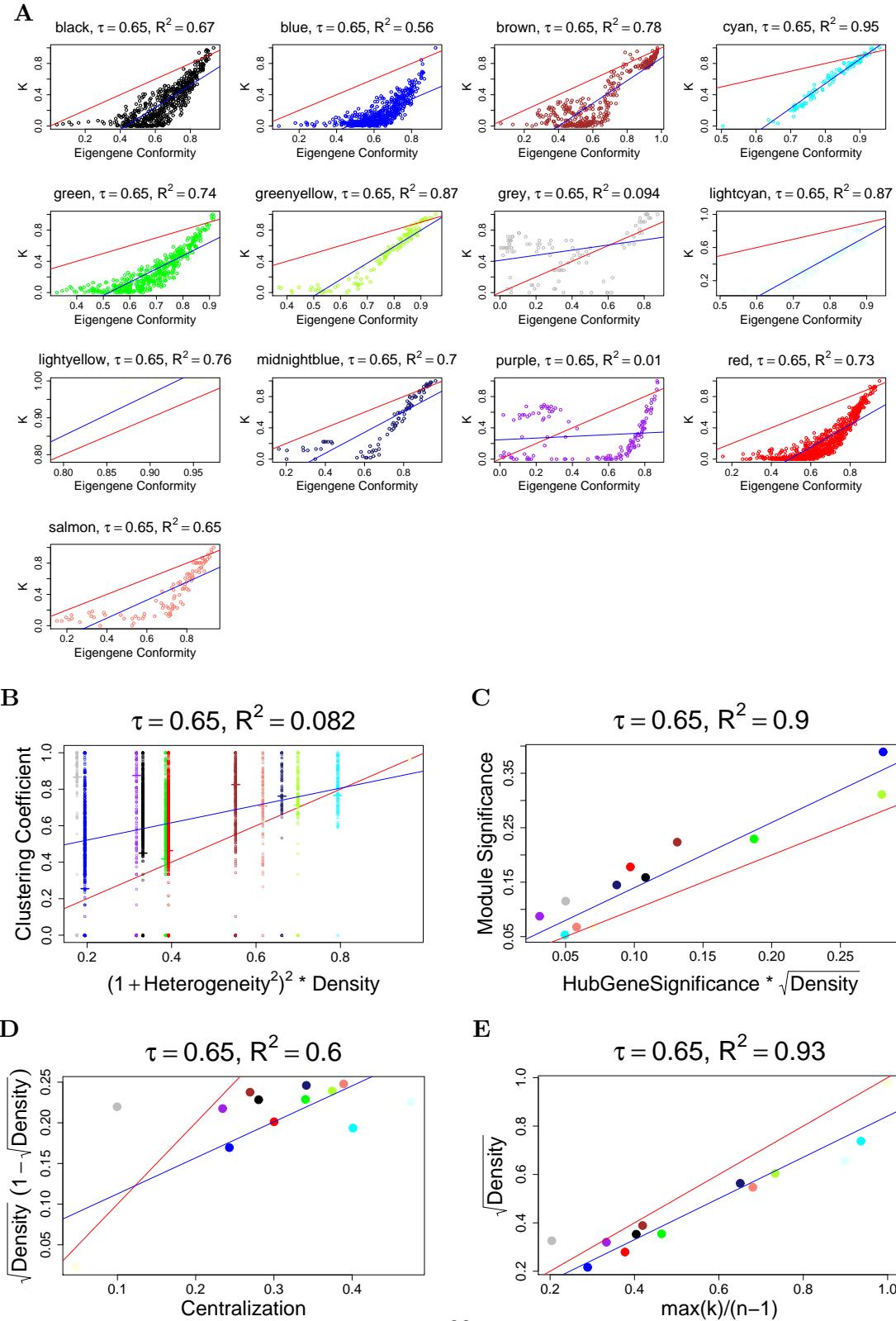


Figure 22: This figure is analogous to Figure 8 in the main article, corresponding to an unweighted network constructed with a hard threshold of $\tau = 0.65$. It illustrates Observation 3 regarding the relationships among network concepts.

11 Unweighted Gene Co-Expression Network Results for $\tau = 0.5$

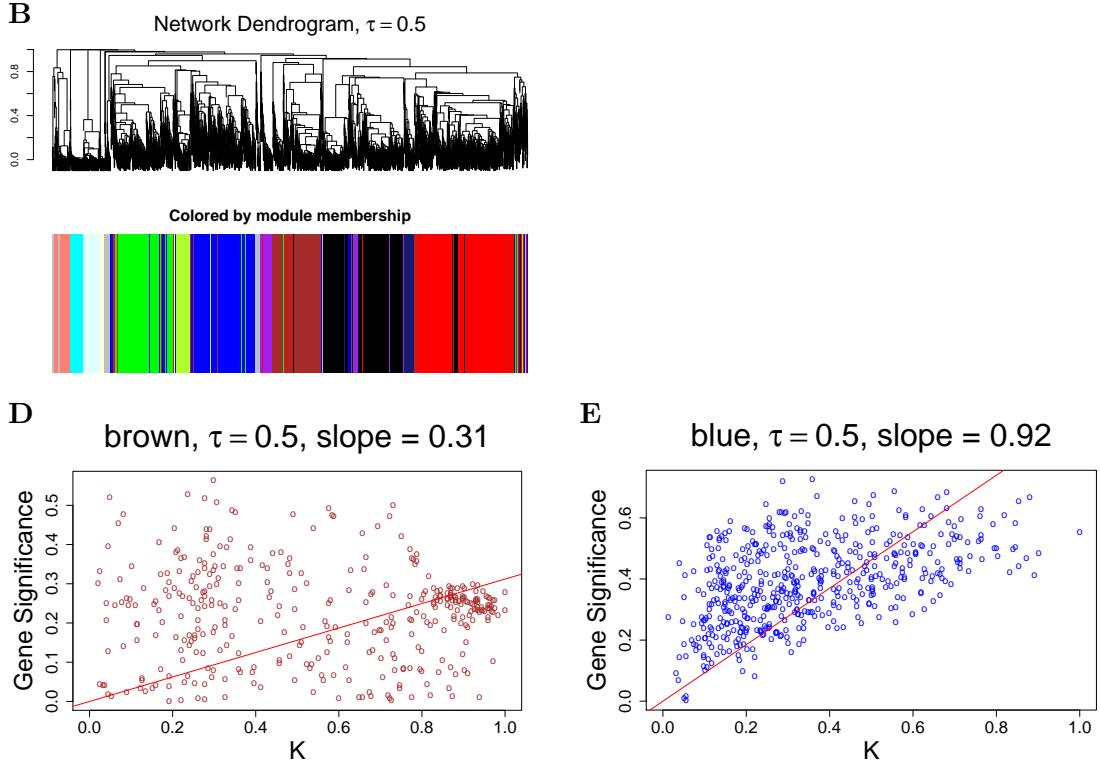


Figure 23: This figure is analogous to Figure 3 in the main article. The only difference is that we use a different dissimilarity for constructing the cluster tree in Figure B. Specifically, we use the topological overlap matrix based dissimilarity corresponding to an unweighted network constructed with $\tau = 0.5$. Figure B depicts the hierarchical cluster tree of genes. Modules correspond to branches of the tree. The branches and module genes are assigned a color as can be seen from the color-bands underneath the tree. Grey denotes genes outside of proper modules. Figure C would be identical to the case of $\beta = 1$, and thus is omitted here. Figures D and E show scatter plots of gene significance GS (y-axis) versus scaled connectivity K (x-axis) in the brown and blue module, respectively. The underlying gene significance is defined with respect to the mouse body weight. The hub gene significance is defined as the slope of the red line, which results from a regression model without an intercept term.

Table 9: Values of network concepts for an unweighted network constructed with a hard threshold of $\tau = 0.5$.

Module	black	blue	brown	cyan	green	greenyellow	grey	lightcyan	lightyellow	midnightblue	purple	red	salmon
Size ($n^{(q)}$)	548	534	366	96	406	121	104	119	34	84	139	772	98
Eigengene Fac. ($EF(X^{(q)})$)	0.898	0.91	0.895	0.99	0.938	0.973	0.369	0.975	0.996	0.93	0.689	0.921	0.913
$VarExplained(E^{(q)})$	0.442	0.401	0.465	0.667	0.478	0.592	0.196	0.64	0.799	0.542	0.348	0.425	0.525
$max(a_{e,i})$	0.932	0.928	0.977	0.957	0.915	0.955	0.874	0.935	0.973	0.963	0.87	0.949	0.933
Density	0.363	0.244	0.369	0.92	0.433	0.662	0.151	0.843	1	0.563	0.317	0.303	0.584
Density E	0.422	0.387	0.422	0.667	0.464	0.578	0.122	0.639	0.82	0.507	0.274	0.406	0.49
Centralization	0.417	0.463	0.303	0.0822	0.401	0.233	0.0838	0.16	0	0.263	0.244	0.467	0.299
Centralization E	0.185	0.192	0.215	0.121	0.161	0.153	0.188	0.114	0.079	0.187	0.186	0.199	0.17
Heterogeneity	0.46	0.578	0.539	0.122	0.454	0.368	0.374	0.173	0	0.402	0.459	0.61	0.361
Heterogeneity E	0.218	0.193	0.323	0.101	0.174	0.18	0.786	0.103	0.0576	0.286	0.527	0.216	0.289
Mean(ClusterCof E)	0.753	0.624	0.771	0.948	0.732	0.87	0.901	0.902	1	0.884	0.781	0.701	0.844
ClusterCof E	0.462	0.416	0.513	0.674	0.492	0.611	0.317	0.647	0.802	0.587	0.444	0.445	0.569
ModuleSignif	0.159	0.389	0.223	0.0531	0.229	0.311	0.115	0.0694	0.069	0.145	0.0875	0.178	0.0676
ModuleSignif E	0.0935	0.387	0.177	0.0167	0.219	0.303	0.0646	0.0388	0.0618	0.0998	0.00798	0.119	0.0399
HubGeneSignif	0.265	0.924	0.312	0.0557	0.384	0.376	0.149	0.0775	0.069	0.153	0.0954	0.314	0.087
HubGeneSignif E	0.134	0.578	0.266	0.0196	0.295	0.382	0.162	0.0456	0.0674	0.136	0.0133	0.177	0.0535
EigengeneSignif $= a_{e,t}^{(q)}$	0.144	0.623	0.272	0.0205	0.322	0.4	0.186	0.0488	0.0693	0.141	0.0153	0.187	0.0573

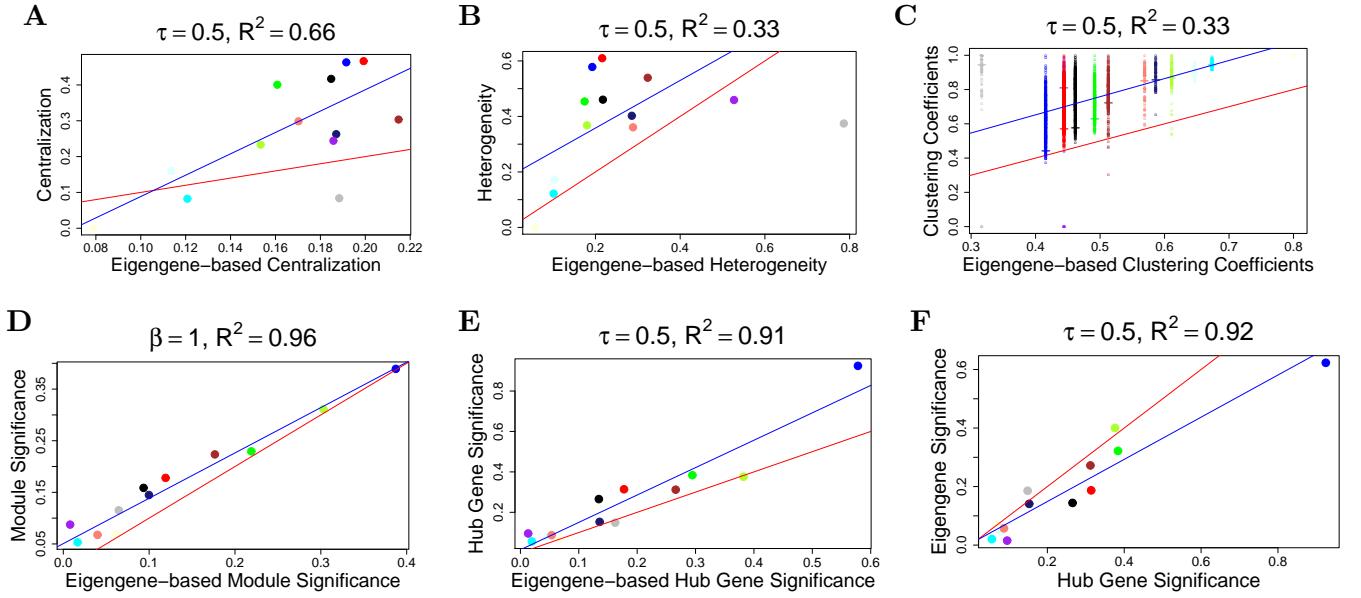


Figure 24: This figure is analogous to Figure 6 in the main article, corresponding to an unweighted network constructed with a hard threshold of $\tau = 0.5$. It illustrates Observation 2 regarding the relationship between network concepts (y-axis) and their eigengene-based analogs (x-axis) in the mouse data. Each point corresponds to a module. Figure A: Centralization (y-axis) versus eigengene-based Centralization_E (x-axis); analogous plots for Figure B: Heterogeneity; Figure C: clustering coefficient; Figure D: module significance; and Figure E: hub gene significance. Figure D is identical to the case of $\beta = 1$. Figure F illustrates the relationship between eigengene significance and hub gene significance. The blue line is the regression line through the points representing proper modules (i.e., the grey, non-module genes are left out). While the red reference line (slope 1, intercept 0) does not always fit well, we observe high squared correlations R^2 between network concepts and their analogs. Since the grey point corresponds to the genes outside properly defined modules, we did not include it in calculations.

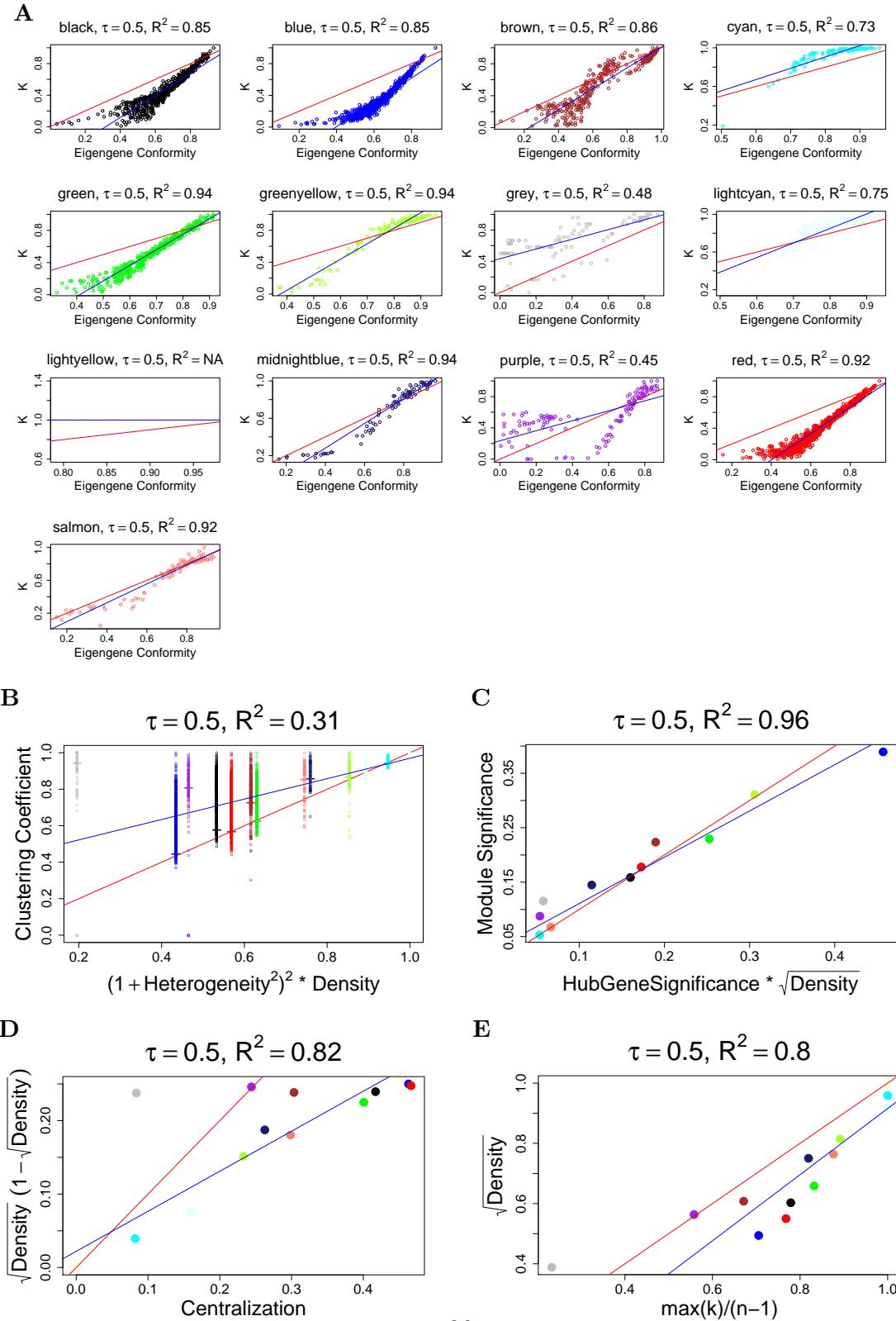


Figure 25: This figure is analogous to Figure 8 in the main article, corresponding to an unweighted network constructed with a hard threshold of $\tau = 0.5$. It illustrates Observation 3 regarding the relationships among network concepts.

References

Ghazalpour, A., Doss, S., Zhang, B., Plaisier, C., Wang, S., Schadt, E.E., Thomas, A., Drake, T.A., Lusis, A.J. and Horvath, S. (2006). Integrating genetics and network analysis to characterize genes related to mouse weight. *PloS Genetics*, **2**(8).