

webPSN
Output Files Guide

Contents

1.	Introduction	p	3
2.	Outputs from single structure PSN analysis	p	4
2.1.	Data files	p	4
2.1.1.	links.csv	p	4
2.1.2.	hubs.csv	p	5
2.1.3.	paths_filter_info.txt	p	5
2.1.4.	paths_summary.csv	p	6
2.1.5.	Path distribution files	p	7
2.1.6.	metapath.csv	p	8
2.2.	Plots	p	9
2.2.1.	Path distribution plots	p	9
2.2.2.	MetaPath 2D	p	9
2.3.	3D Script Files	p	10
2.3.1.	links.pml, links.vmd	p	10
2.3.2.	hubs.pml, hubs.vmd	p	11
2.3.3.	comms.pml, comms.vmd	p	11
2.3.4.	metapath.pml, metapath.vmd	p	11
3.	Outputs from network difference calculations	p	12
3.1.	Data files	p	12
3.1.1.	align.csv, align.fasta	p	12
3.1.2.	info.csv	p	13
3.1.3.	links.csv	p	14
3.1.4.	hubs.csv	p	14
3.1.5.	paths_filter_info.txt	p	15
3.1.6.	mpdiff_info.csv	p	15
3.1.7.	mpdiff_tab.csv	p	16
3.2.	Plots	p	17
3.2.1.	Network Difference plots	p	17
3.2.2.	Difference path distribution plots	p	17
3.2.3.	mpdiff2d.png	p	18
3.3.	3D Script Files	p	18
4.	Outputs from consensus network calculations	p	18
5.	Contact information and how to cite	p	19

1. Introduction.

This document describes the output files produced by webPSN server present in the downloadable compressed zip archives.

With the only two exceptions of submitted pdb files, which retain their original unmodified names, and pdb files download from the Protein Data Bank, which will be named by the server after their code (e.g. 1U19.pdb), the name of all other output files will starts with the common prefix *webpsn_output_jobID_*, where **ID** is a unique alphanumeric code assigned by the server to your calculation (e.g. *webpsn_output_jobB72dKV_paths_summary.csv*).

To download again the compressed archive with all your results you can follow the link in the email notification we sent to you when your analysis is complete (if you provided your email address during submission) or copying the unique job id in the specific form of the [Retrieve](#) section of webPSN server.

For the sake of clarity and readability in the following pages all output files will be listed without the common prefix (i.e. *paths_summary.csv* instead of *webpsn_output_jobB72dKV_paths_summary.csv*).

Almost all data files present in the zip archive are in *csv* format, that can be opened ed edited using common spreadsheet software (e.g. LibreOffice Calc, MS Excel and Google Sheets). Other data files are simple text file that can be opened with any text editor (e.g. Notepad, TextEdit). Among output files there are also several plots in *png* format that can be visualized with any image viewers and several script files that can be used to produce 3D representation of your network. The latter are present for both, PyMol and VMD molecular visualization software. Please refer to their respective manuals for more information about how to use the scripts.

2. Outputs from single structure PSN analysis

This section provides a detailed description of the output files produced during the PSN analysis of a single structure and present in the downloadable compressed zip archive.

2.1. Data Files

The following sections describe the content of all produced data files. As stated above almost all are *csv* files, that can be opened with a spreadsheet software, while those with the *txt* extension are simple text files.

2.1.1. links.csv

This csv file contains a table that lists all the links present in the network and the columns arrangement is almost identical to that present in the “*Links*” table of the “*Viewer & Tables*” section of the result page. The columns in this table have the following meanings:

- **N**: progressive link id, irrelevant for the user.
- **Node1, Node2**: the two interacting nodes that form the link.
- **OutValue1, OutValue2**: these columns report the values present in the user submitted external values file (if provided during submission).
- **Freq**: always 0 or 100, irrelevant for the user.
- **Force**: the interaction strength between the two nodes.
- **IsNode1Hub?, IsNode2Hub?**: "Yes" if the corresponding node is a hub (i.e. the node has more than 3 links), otherwise "No".
- **Clust**: the cluster of nodes the link belong to, see the [online documentation page](#) for more information about this topic.
- **Comm**: the id of the community the link belong to, otherwise 0.

2.1.2. hubs.csv

This csv file contains the table that lists all the hubs present in the network and the columns arrangement is almost identical to that present in the “Hubs” table of the “Viewer & Tables” section of the result page. The columns in this table have the following meanings:

- **N**: progressive hub id, irrelevant for the user.
- **Hub**: the hub being considered.
- **OutValue**: this columns reports the values present in the user submitted external values file (if provided during submission).
- **Freq**: always 0 or 100, irrelevant for the user.
- **Force**: the average interaction strength of the links of the corresponding hub.
- **Clust**: the cluster of nodes the hub belong to, see the online documentation for more information about this topic.
- **Comm**: the id of the community the hub belong to, otherwise 0.

2.1.3. paths_filter_info.txt

This is a small text file which serves as a remainder of the last path filtering applied. The first three lines of the file are a header, the unique job id, and the pdb file name. The remaining lines depend on the applied path filter and may contains the following lines:

- **Apical**: the apical node from which all paths start from or end to.
- **Pair**: the nodes from which all paths start from and end to.
- **Pass**: the node(s) all paths pass through.

This file and all other files which include the text *_filtered* in the file name are present only if a path filtering is applied **before** the download.

2.1.4. paths_summary.csv

This csv file contains a summary table with some information and statistics about the shortest paths of the PSN analysis. This file is always present in the zip archive and, if a path filtering is applied before the download, another file named *paths_summary_filtered.csv* will be present as well. Both files share the same contents and organization, the former refers to the global pool of paths (i.e. without applied path filters) while the data present in the latter refers to the filtered pool of paths. These two tables report almost the same values present in "Path Summary" and "Filtered Path Summary" tables present in the *Summary Tables* sections of the result page.

After the header there are 23 rows which report the following information:

- **NetName**: always PSN or consensus, irrelevant for the user.
- **MinCorr**: always 0.7. This is the minimum ENM correlation that at least one node in each path must have with the first or last node in the path. This value is fixed and is set so high to ensure a solid correlation of the atomic fluctuation among the nodes in the paths.
- **NumOfPaths**: total number of paths in the global/filtered path pool.
- **Min-, Max-, Avg-, StD- PathsLength**: the shortest, the longest, the average and the corresponding standard deviation of the length of paths expressed as number of nodes in the path.
- **Min-, Max-, Avg-, StD- PathsForce**: the lowest, the highest, the average and the corresponding standard deviation of the average interaction strength among the links present in the global/filtered path pool.
- **Min-, Max-, Avg-, StD- PathsCorr**: the lowest, the highest, the average and the corresponding standard deviation of the average correlation between the each node and the first and last nodes in each path in the global/filtered path pool.
- **Min-, Max-, Avg-, StD- PathsScore**: the lowest, the highest, the average and the corresponding standard deviation of the percentage of

nodes with a correlation ≥ 0.7 with the first and/or the last node in the path, present in the global/filtered path pool.

- **Min-, Max-, Avg-, StD- PathsHubs%:** the lowest, the highest, the average and the corresponding standard deviation of the percentage of hub nodes present in the global/filtered path pool.

2.1.5. Path distribution files

These are a set of 5 csv files all organized in the same manner and that store the same type of information and for this reason are all detailed in this section. As for the previous files, if a path filters is applied, an additional set of 5 *_dist_filtered.csv* companion files will be present in the downloaded archive.

These tables report the distribution of some descriptors in the global/filtered pool of paths and are used to generate the corresponding plots visible on the webPSN server in the “*Path Graphs*” and “*Filtered Path Graphs*” sections of the result page and also present in the downloadable zip archive as *png* files.

All tables share the same structure composed by 3 columns: the first, specific for each file, reports an increasing value of the specific descriptor, while the remaining two columns report the number and the percentage of the global/filtered pool of paths with the corresponding descriptor value. The paths distribution files are:

- **paths_length_dist.csv:** the distribution of the number of nodes in each path of the global/filtered path pool.
- **paths_force_dist.csv:** the distribution of the average interaction strength of the links in each path of the global/filtered path pool.
- **paths_corr_dist.csv:** the distribution of the average correlation between the each node and the first and last nodes in each path of the global/filtered path pool.

- **paths_corrfract_dist.csv**: the distribution of the percentage of nodes with a correlation with the first and/or the last node in each path present in the global/filtered path pool.
- **paths_hubs_dist.csv**: the distribution of the percentage of hubs in each path present in the global/filtered path pool.

2.1.6. metapath.csv

This csv file contains a detailed report of the links and nodes that contribute to the global/filtered metapath. As for the *paths_summary.csv* file detailed above, a version of this file, called *metapath_filtered.csv*, is present in the zip archive if a path filters was applied before the download. This table is almost identical to those called “Global/Filtered MetaPath” in the “Viewer & Tables” section of the result page. The columns in this table have the following meanings:

- **N**: progressive link id, irrelevant for the user.
- **Node1, Node2**: the two interacting nodes that form the metapath link.
- **OutVal1, OutVal2**: these columns report the values present in the user submitted external values file (if provided during submission).
- **LinkFreq**: always 0 or 100, irrelevant for the user.
- **LinkForce**: the interaction strength between the two nodes.
- **LinkRec**: the relative recurrence of the link in the filtered pool of shortest paths.
- **Node1Rec, Node2Rec**: the average recurrence of the links of each nodes.
- **IsNode1Hub, IsNode2Hub**: "Yes" if the corresponding node has more than 3 links, otherwise "No".

2.2. Plots









This section briefly describes the plot files present in the downloadable zip archive. As for other output files these plots are available for the global paths pool and, if a path filter was applied, for the resulting filtered path pool.

2.2.1. Path distribution plots

These plots are produced using the data present in the path distribution data files detailed in [section 2.1.5](#). For more details about these plots refer to the description given about the corresponding data files.

2.2.2. mpaths2d.png

This is a 2D representation of the global/filtered metapath also available on the webPSN server. Regular nodes are represented as circles while hubs are drawn as pentagons. Links are represented as lines connecting two nodes. The colors of nodes and links represent the relative recurrence of nodes and links in the global/filtered path pool and are the same used in the corresponding 3D representations present on the webPSN server and in the 3D script files. The following table illustrates the level of recurrence associated to each color:

Color	Recurrence
	$25 \leq r \leq 30$
	$30 < r \leq 40$
	$40 < r \leq 50$
	$50 < r \leq 60$
	$60 < r \leq 70$
	$70 < r \leq 80$
	$80 < r \leq 90$
	$90 < r \leq 100$

2.3. 3D Script Files











This section lists the 3D script files present in the downloadable zip archive. These scripts map the network on the protein/nucleic acid structure used for the calculations and are equivalent to those present on the webPSN server in the “Viewer & Tables” section of the result page.

In these representations each node is represented as a sphere centered on the Ca carbon atom of standard amino acids, on the N1 atom of standard nucleotides and on the atom nearest to the geometric center for all other molecules present in the analyzed structure. The server produces scripts for both PyMol and VMD molecular visualization softwares. The two versions of the same script are equivalent and use the same colors and styles.

2.3.1. links.pml, links.vmd

These scripts reproduce all the links present in the calculated network and their nodes. Links are colored according to their interaction strength and nodes according to the average interaction strength of their links. The following table illustrates the value of the interaction strength associated to each color:

Color Interaction Strength










	$0 < i.s. \leq 1$
	$1 < i.s. \leq 2$
	$2 < i.s. \leq 3$
	$3 < i.s. \leq 4$
	$4 < i.s. \leq 5$
	$5 < i.s. \leq 6$
	$6 < i.s. \leq 7$
	$7 < i.s. \leq 8$
	$8 < i.s. \leq 9$
	$9 < i.s. \leq 10$

2.3.2. hubs.pml, hubs.vmd

These scripts show the hubs of the network colored according to the average interaction strength of their links and using the same colors detailed in the previous section.

2.3.3. comms.pml, comms.vmd

These scripts display communities of nodes identified in the calculated network. Each community is represented with a unique color and nodes and links belonging to the same community share the same colors. The first nine most populous communities are colored as follow:

Color	Community
	1 st most populous community
	2 nd most populous community
	3 rd most populous community
	4 th most populous community
	5 th most populous community
	6 th most populous community
	7 th most populous community
	8 th most populous community
	9 th most populous community

2.3.4. metapath.pml, metapath.vmd

These scripts represent methpath calculated on the global path pool. If a path filter was also applied, the corresponding filtered version will be present in

the zip archive (named `metapath_filtered.pml`, `metapath_filtered.vmd`). Nodes and links are colored according to their recurrences and the colors used are the same detailed for the 2D version in [section 2.2.2](#).

3. Outputs from network difference calculations.

This section provides a detailed description of the output files produced in a network difference analysis divided by file type. Most of these files share some similarities with those presented in the previous sections.

3.1. Data Files

The following sections describe the content of all data files produced. As stated above almost all are csv files, which can be opened with a spreadsheet software, while those with the `txt` extension are simple text files.

3.1.1. `align.csv`, `align.fasta`

The file `align.csv` stores the same table present at the bottom of the “*Single Jobs and Labels*” section of the result page. This table reports the label assigned to each node in the analyzed networks (column **Lab**), the percentage of networks in which that node is present (column **PresentIn**) and the corresponding residue in each pdb file, reported as:

Chain:Segment:ResTypeResNum.

The associated `align.fasta` text file reports the same multiple sequence alignment present in the “*Single Jobs and Labels*” section of the result page.

3.1.2. info.csv

This csv file contains a three columns table almost identical to that present in the “*Summary Tables*” section of the result page. The first column lists a some network properties and the following two columns the corresponding values in the two analyzed networks. After the header line there are 18 rows with the following meaning:

- **Freq:** always zero, irrelevant for the user.
- **I_{min}:** a network specific value that indicates the lowest interaction strength needed to link two nodes. Please refer to the [theory section](#) of the online documentation for more details about this value.
- **LNodes:** the number nodes with at least one link.
- **Links:** the number of links.
- **Hubs:** the number of hubs.
- **HLinks:** the number of links mediated by hubs.
- **SpecLinks, SpecLinks%:** the number of links, and the corresponding percentage, present only in one of the two analyzed networks.
- **SharedLinks, SharedLinks%:** the number of links, and the corresponding percentage, shared by both networks.
- **SpecNodes, SpecNodes%:** the number of nodes with at least one link, and the corresponding percentage, present only in one of the two analyzed networks.
- **SharedNodes, SharedNodes%:** the number of nodes with at least one link, and the corresponding percentage, shared by both networks.
- **SpecHubs, SpecHubs%:** the number of hubs, and the corresponding percentage, present only in one of the two analyzed networks.
- **SharedHubs, SharedHubs%:** the number of hubs, and the corresponding percentage, shared by both networks.

3.1.3. links.csv

This csv table is arranged similarly to that present in the “*Links*” table of the “*Viewer & Tables*” section of the result page. The columns in this table have the following meanings:

- **N**: progressive link id, irrelevant for the user.
- **Node1, Node2**: the two interacting nodes.
- **Owner**: the name of the network the link belong to or "Shared" if the corresponding link is present in both networks.
- **Freq1, Freq2**: always 0 or 100, irrelevant for the user.
- **Force1, Force2**: the interaction strength between the two nodes or 0 if the link is not present.
- **OutVal1, OutVal2**: these columns report the values present in the user submitted external value file (if provided during submission).
- **IsNode1HubInNet1, IsNode1HubInNet2**: "Yes" if the corresponding node has more than 3 links in the first network, otherwise "No".
- **IsNode2HubInNet1, IsNode2HubInNet2**: "Yes" if the corresponding node has more than 3 links in the second network, otherwise "No".

3.1.4. hubs.csv

This csv table reproduces that present in the “*Hubs*” table of the “*Viewer & Tables*” section of the result page. The columns in this table have the following meanings:

- **N**: progressive link id, irrelevant for the user.
- **Hub**: the hub being considered.
- **Owner**: the name of the network the hub belong to or "Shared" if is present in both networks.
- **Degree1, Degree2**: the number of links the considered hub has in the first and second network.

- **Freq1, Freq2:** always 0 or 100, irrelevant for the user.
- **Force1, Force2:** the average interaction strength of the links of considered hub in the two networks.

3.1.5. paths_filter_info.txt

This small text file serves as reminder of the last applied path filtering. For more details read the description given in [section 2.1.3](#).

3.1.6. mpdiff_info.csv

This csv table contains a comparison summary of the metapaths calculated on the two structure networks. This file is always present in the zip archive and, if a path filtering is applied before the download, another file named *mpdiff_info_filtered.csv* will be present as well. Both files share the same contents, the former refers to the global pool of paths (i.e. without path filters) while the data present in the latter refers to the filtered pool of paths. The first column lists a series of network properties and the following two columns the corresponding values in the two analyzed networks. After the header line there are 14 rows with the following meaning:

- **MinFreq:** always 0, irrelevant for the user.
- **MinCorr:** always 0.7. This is the minimum ENM correlation that at least one node in each path must have with the first or last node in the path. This value is fixed and is set so high to ensure a solid correlation of the atomic fluctuation among the nodes in the paths.
- **MinRec:** always 10. This is the minimum relative recurrence in the path pool needed by a link to be represented in a metapath.
- **TotPaths:** the total number of global/filtered shortest paths.
- **MPLinks:** the number of links in the global/filtered metapaths.
- **MPNodes:** the number of nodes in the global/filtered metapaths.

- **SpecLinks, SpecLinks%:** the number of metapath links, and the corresponding percentage, present only in one of the two analyzed networks.
- **SharedLinks, SharedLinks%:** the number of metapath links, and the corresponding percentage, shared by both networks.
- **SpecNodes, SpecNodes%:** the number of metapath nodes, and the corresponding percentage, present only in one of the two analyzed networks.
- **SharedNodes, SharedNodes%:** the number of metapath nodes, and the corresponding percentage, shared by both networks.

3.1.7. mpdiff_tab.csv




As for the previous table, this csv file is always present in the downloadable zip archive and is associated to another file, called *mpdiff_tab_filtered.csv*, if a path filter is applied before download. This table is almost identical to those called “Global/Filtered MetaPath” in the “Viewer & Tables” section of the result page. The columns in this table have the following meanings:

- **N:** progressive metapath link id, irrelevant for the user.
- **Node1, Node2:** the two interacting nodes that form the metapath link.
- **Owner:** the name of the network the link belongs to or "Shared" if the corresponding link is present in both metapaths.
- **Rec1, Rec2:** the link recurrence in both metapaths.
- **Freq1, Freq2:** always 0 or 100, irrelevant for the user.
- **Force1, Force2:** the interaction strength between the two nodes or 0 if the link is not present.
- **OutVal1, OutVal2:** these columns report the values present in the user submitted external values file (if provided during submission).
- **IsNode1HubInNet1, IsNode1HubInNet2:** "Yes" if the corresponding node has more than 3 links in the first network, otherwise "No".
- **IsNode2HubInNet1, IsNode2HubInNet2:** "Yes" if the corresponding node has more than 3 links in the second network, otherwise "No".

3.2. Plots

This section briefly describes the plots produced in a networks difference analysis. As for other output files some of these plots are available for the global path pool and, if a path filter is applied, for the resulting filtered pool of paths. The colors used in all plots are the same of those used on the web server and have the following meaning:

Color Meaning

	Present only in the 1 st network
	Shared by both networks
	Present only in the 2 nd network

3.2.1. Network Difference plots

This section describes three png files named: *spec_comm_lnodes.png*, *spec_comm_links.png* and *spec_comm_hubs.png*. These plots are produced using the data present in the *info.csv* file detailed in [section 3.1.2.](#) and graphically compare the number of linked nodes (i.e. nodes with at least one link), the number of links and the number of hubs (i.e. the number of nodes with at least 4 links) in the two analyzed networks. These network components are expressed in terms of specific vs shared and the colors used in these plots are the same detailed above.

3.2.2. Difference path distribution plots

These plots compare the distribution of some descriptors in the global/filtered pool of paths between the two analyzed networks. For more details about these plots refer to the description given in [section 2.1.5.](#)

3.2.3. **mpdiff2d.png**

This plots, and the corresponding filtered version *mpdiff2d_filtered.png*, compares, with a 2D representation, nodes and links participating in the global/filtered metapath of the two analyzed networks. For a description of this plot see [section 2.2.2](#). The colors used are the same described above.

3.3. 3D Script Files

This section briefly lists the 3D script files present among the network difference output files. These scripts compare links, hubs, global metapaths and filtered metapaths (if a path filter was applied) and are named: *linksdiff*, *hubsdiff* *mpdiff* *mpdiff_filtered*, respectively. There are two version for each 3D script, for both PyMol and VMD molecular visualization softwares and all use the same colors detailed in [section 3.2](#). For more information about these files refer to the corresponding section detailing the 3D scripts produced in the single structure PSN analysis in [section 2.3](#).

4. Outputs from consensus network calculations

The result of a consensus analysis is a new average network generated from a pool of single structure networks. All data files, plots and 3D scripts produced by a consensus analysis shares the same names, format and organization of those produced in a single network analysis. Please refer to those sections for more details.

5. Contact information and how to cite

Thank you for using webPSN, we really appreciate it.

For any comments or questions about webPSN server, please feel free to contact us at the following email addresses:

- francesca.fanelli@unimore.it
- angelo.felline@unimore.it

Please, remember to cite the following paper in all published works which utilizes this web server:

*Felline A., Seeber M. Fanelli F. - “**webPSN v2.0: a webserver to infer fingerprints of structural communication in biomacromolecules**” - Nucleic Acids Res, Web Server Issue, 19 May 2020*
<https://doi.org/10.1093/nar/gkaa397>