

Summary

This document describes the functionalities and processing options of the Conefor command line version for directed graphs, i.e. for networks with asymmetric connectivity. The majority of options and modes of use described in this document are similar to, if not the same as, those of the standard (non-directed) graphical user interface (GUI) version of Conefor and of the standard (non-directed) command line version of Conefor. Therefore, more detailed specifications and explanations can be found in the user's manuals for these other previous non-directed versions as available at www.conefor.org, as well as in the key papers on the connectivity metrics and methods available in that website. In particular, examples of how to use the command line syntax to run certain Conefor analyses, as well as guidance on how to call from R the Conefor command line (either for directed or non-directed graphs, since the procedure is the same) can be found in the files that are downloaded with the non-directed Conefor command line as available in that website.

Since the usage of this directed version is very similar to that of the standard (non-directed) command line Conefor, if you have previously used the standard command line Conefor, you should have no problems in readily using this directed version as well.

A summary of the most notable differences in the directed version of Conefor command line compared to the non-directed one is the following:

- The connection file for the directed version can have the double number of lines than the non-directed connection file. In the non-directed version, the connection file for a network with n nodes can have up to $n(n-1)/2$ lines; this is the number of lines in a full connection file, i.e. a connection file in which all pairs of nodes are listed. In the directed version, however, the connection file can have up to $n(n-1)$ lines (and exactly that number for a full non-directed connection file). This is because, in directed networks, the probability of moving from a node i to a node j can be different from the probability of moving from node j to node i ; therefore, each pair of nodes has to be listed twice, in both directions. As in the directed version, if a partial connection file is used in which not all pairs and directions are listed, a zero probability of direct movement between those missing nodes, or along those missing directions, will be assumed.
- The second difference is that this directed Conefor version only includes probabilistic indices: Probability of Connectivity (PC), Flux (F), Area-Weighted Flux (AWF), and BC(PC), i.e. the Betweenness Centrality improved/generalized to acquire some of the desirable properties of PC. These indices use a network in which links are weighted by the probability of movement/dispersal. This directed Conefor version does not currently include binary indices such as IIC (using networks with unweighted links), although they may be incorporated in a future version.
- Link-level analyses (link removal, link improvement, link change) are not available in this version for directed networks, although they may be included in a future compilation.
- If the PC index has been selected for calculation, the file with the results of the calculated overall indices will also include the values of the three fractions of the PC index (PCintra, PCdirect, PCstep; expressed as a percentage of the total PC value) that are described in Appendix S5 in Saura et al. (2014) (see reference below in this section). Note that these three fractions correspond to the overall PC value for the entire landscape/network, and that they are different from the node-level fractions of dPC (intra, flux, connector) that were already implemented in previous versions of Conefor and that are provided for each node in the node importance file. The last paragraph of Appendix S5 in Saura et al. (2014) provides additional information on the differences between the overall landscape and the node-level fractions of PC. These overall-landscape PC fractions will also be incorporated in the non-directed version of Conefor soon.

Having noted these differences, Conefor for directed networks can be used similarly as the standard non-directed command line version already available before. In any case, a description of all options and parameters, and of some other smaller differences, is provided in the next section below.

To date, Conefor for directed networks has been used in the following studies involving different types of asymmetric connectivity, namely spatio-temporal connectivity (Martensen, Saura and Fortin 2017), waterbird migration (Merken et al. 2015), riverscapes (Rincón et al. 2017) and asymmetric dispersal and colonization likelihood because of differences in the number of dispersing individuals in the nodes (Saura, Bodin and Fortin 2014):

- Martensen, A., Saura, S., Fortín, M.J. 2017. Spatio-temporal connectivity: assessing the amount of reachable habitat in dynamic landscapes. *Methods in Ecology and Evolution*. <http://dx.doi.org/10.1111/2041-210X.12799>
- Rincón, G., Solana-Gutiérrez, J., Alonso, C., Saura, S., García de Jalón, D. 2017. Longitudinal connectivity loss in a riverine network: accounting for the likelihood of upstream and downstream movement across dams. *Aquatic Sciences* 79: 573-585. <http://dx.doi.org/10.1007/s00027-017-0518-3>
- Merken, R., Deboelpaep, E., Teunen, J., Saura S., Koedam, N. 2015. Wetland suitability and connectivity for trans-Saharan migratory waterbirds. *PLoS ONE* 10(8): e0135445. <http://dx.doi.org/10.1371/journal.pone.0135445>
- Saura, S., Bodin, Ö., Fortin, M.J. 2014. Stepping stones are crucial for species' long-distance dispersal and range expansion through habitat networks. *Journal of Applied Ecology* 51: 171-182. <http://dx.doi.org/10.1111/1365-2664.12179>

Many other cases or systems involving asymmetric connectivity could be examined in the future. We hope that this Conefor version will be of help to support your analyses in other (similar or very different) research projects or conservation applications in which directed networks are involved.

Please use the following suggested format to cite this Conefor version for directed networks: "analyses were performed using a version of the Conefor software (Saura and Torné 2009) recently updated for directed networks and available at www.conefor.org".

The user assumes all responsibility for the accuracy and suitability of this program for any application.

Description of the syntax, options and parameters in the directed version of Conefor command line

If in the command line you simply write *conefor.exe*, or *conefor.exe -help*, you will get, in addition to the information on the Conefor version, a list of the parameters that can be used to run Conefor for directed graphs, which is as follows:

```
conefor.exe -nodeFile nfile -conFile cfile [ -* ] [ -t dist | prob [ all | notall ] ]  
[ -confProb distance prob [ -PC [ -BCPC ] ] [ -F ] [ -AWF ] ]  
[ -double | -simple ] [ onlyoverall ]  
[ -wprobdir ] [ -wprobmax ] [ -noout ]
```

The arguments between brackets [] are optional

The arguments separated by | indicate that only one of them should be chosen; default values are noted below.

Next, we describe each of these parameters.

-nodeFile nfile

nfile specifies the path and file name for the node file. If only the name of the file is specified, it is understood that the file is located in the same folder as *conefor.exe*. The name of the file and path should not contain any spaces.

When the option -* is used (see further below a description of this option), nfile indicates the starting characters of all the node files to be processed in a single run.

-conFile cfile

cfile specifies the path and file name for the node file. If only the name of the file is specified, it is understood that the file is located in the same folder as *conefor.exe*. The name of the file and path should not contain any spaces.

When the option `-*` is used (see further below a description of this option), `cfile` indicates the starting characters of all the connection files to be processed in a single run.

-t dist|prob [all|notall]

The first part indicates the type of connection file:

- `dist`: The connection file is a distance file (default option)
- `prob`: The connection file is a probability file

Note that the `adj` option that is available for the standard non-directed Conefor is not here available in this directed Conefor version. This is because the current Conefor command line for directed graphs only includes probabilistic indices. If you are interested in analysing networks with unweighted links (only 1s or 0s in the connection file), there is the possibility to do so by selecting the `prob` type of connection file; in this case the analysis will be conducted using probabilistic graphs and probabilistic indices in the particular case in which the link probabilities correspond only to 1s or 0s.

The next option indicates if the connections between every pair of nodes, or just a part of them, are reported in the connection file:

- `all`: all the pairs of nodes are listed (only once) in the connections file (also called a full connection file). In such a full connection file for a directed network with n nodes, the number of lines should be $n(n-1)$; otherwise, Conefor will give an error message and will not proceed with the analysis.
- `notall` (default option): only some pairs of nodes are listed in the connection file (also called a partial connection file). The pairs of nodes and/or directions that are not included in the connections file are assumed to have no link (zero probability of direct dispersal or infinite distance between them). Under this option, Conefor will run the analysis for a connection file with less than $n(n-1)$ lines (assuming no links in the missing cases). This is the default (the option that will be used if none of the two is specified).

The connection file for this directed version of Conefor has the same three columns as the non-directed one (see the general Conefor manual for details), but now the directionality of the connections is given as follows:

- the first column indicates the ID of the source node i
- the second column indicates the ID of the destination node j
- the third column gives the probability of dispersal (for option `prob`) or the distance (for option `dist`) from node i to node j . The probability of dispersal (or the distance) from node j to node i , i.e. in the opposite direction, is to be provided in a different line.

For example, in a hypothetical case involving nodes 23 and 57, the information in the connection file regarding these nodes could be, assuming that probabilities (`prob` option) are provided as an input (the same would apply if distances are used for the connection file):

23	57	0.82
57	23	0.34

If one of the two directions is omitted (under the `notall` option, partial connection file) in the example above, we get:

23	57	0.82
----	----	------

Then it will be understood that movement from node 57 (source) to node 23 (destination) is not possible (and that movement in the other direction is quite likely, with a probability of 0.82). The previous line will be in fact giving the same result as the following two lines:

23	57	0.82
57	23	0

Obviously, that this directed version allows for asymmetric connections does not imply that all pairs necessarily need to have different probabilities in both directions, e.g. the following case (involving the connection between hypothetical nodes 145 and 623) is also perfectly possible for one or many pairs in the connection file:

145	623	0.45
623	145	0.45

When using the directed Conefor version, it is even possible to provide as an input a connection file in which all pairs are symmetric (same probability or distance in the two directions for all pairs of nodes). Such case will be processed correctly by the Conefor version for directed graphs, giving the same result as if the non-directed Conefor version was used, although in such case it would be easier (and faster) to use the non-directed Conefor instead. Note that for the same number of nodes (patches) in the network, the directed version will need somewhat more time to complete the calculations than the standard non-directed Conefor, because the connections have to be checked and calculated in the two directions.

Note that if you use a standard non-directed file, as the one needed for the non-directed Conefor, as an input for the directed Conefor version and specifying the “all” option, then you will get an error because half of the expected lines are missing (as noted above, a full connection file for the non-directed version has a maximum of $n(n-1)/2$ lines, while a full connection file for the directed version of Conefor needs to have $n(n-1)$ lines). If you use a standard non-directed file, as the one needed for the non-directed Conefor, as an input for the directed Conefor but you specify the “not all” option, then you will not get any error message (the “not all” option allows for not pairs/directions to be included), but note that you will get different results than with the non-directed Conefor, because all the “missing” connections in the connection file (one of the directions) will be understood as having infinite distance or null probability of dispersal. The only way to get, with the directed Conefor version, the same results as with the non-directed one, is to duplicate the rows in the non-directed connection file but changing the order of the nodes in the two first columns of the duplicated part.

-*

When this option is specified, Conefor command line will process all the node and connection files that (i) are located in the same folder as coneфор.exe, and (ii) have a name starting with the text indicated by *nfile* (for the node files) and by *cfile* (for the connection files). Additionally, Conefor will automatically generate a prefix for writing the result files corresponding to each calculation; such prefix is the part of the node file name after the characters specified by *nfile* (since all the calculations in this -* mode correspond to node files with name starting with *nfile*).

To use this option there must be, within the same folder where coneфор.exe is located, the same number of node files with name starting with *nfile* as the number of connection files with name starting with *cfile*. In addition, it is necessary that there are pairs of files (one node file and one connection file) with similar names, i.e. with exactly the same name after the characters corresponding to *nfile* and *cfile* (otherwise Conefor will report an error), apart from obviously corresponding to the same landscape/graph.

For example, if *nfile* is *nodes_* and *cfile* is *distances_*, the following list of files (all located in the same folder as coneфор.exe) is correct:

```
nodes_habitat1.txt
nodes_habitat2.txt
nodes_landscape_north.txt
distances_habitat1.txt
distances_habitat2.txt
distances_landscape_north.txt
```

If, in this example, *nfile* was *nodes_h* and *cfile* was *distances_h*, processing would be completed as well with no errors, although only two of the three pairs of files listed above would be processed.

The following set of files will be however incorrect if *nfile* is *nodes_* and *cfile* is *distances_*, since there are three node files with name starting with *nfile* but only two connection files with name starting with *cfile*:

```
nodes_habitat1.txt
nodes_habitat2.txt
nodes_landscape_north.txt
distances_habitat1.txt
distances_habitat2.txt
```

The following set of files is also incorrect (if *nfile* is *nodes_* and *cfile* is *distances_*), since the names of the files *nodes_landscape_north.txt* and *distances_landscape_N.txt* do not match. Assuming that the content of both files is correct (i.e., they correspond to the same landscape/graph), the problem could be solved by renaming the connection file as *distances_landscape_north.txt* or by renaming the node file as *nodes_landscape_N.txt*:

```
nodes_habitat1.txt
nodes_habitat2.txt
```

nodes_landscape_north.txt
distances_habitat1.txt
distances_habitat2.txt
distances_landscape_N.txt

[-confProb distance prob [-PC [-BCPC]] [-F] [-AWF]]

The first part of this set of parameters (*-confProb distance prob*) should be used only when the input connection file is a distance file (*-t dist* above). In such case, the parameters *distance* and *prob* correspond to the distance and probability values that will be used to convert the distance values into probability values, as needed by the probabilistic indices calculated by Conefor. This pair of *distance* and *prob* values will determine the decay rate of a negative exponential function of distance (either Euclidean, effective, or of other form) that will give, for each link, the p_{ij} values (probabilities of direct dispersal) used by the probabilistic indices.

Example: `-confProb 10000 0.5`

If the connection file is a probability file (*-t prob*), none of this parameters needs to be specified, and only the list of probabilistic indices to be computed (see next) should be provided directly.

After the specifications above, one or more of the probabilistic indices to be computed need to be specified (PC, BCPC, F, AWF):

- PC: The PC index will be calculated.
- BCPC: The BC(PC) index will be calculated. Note that this index can only be calculated if PC is also selected.
- F: The F index will be calculated.
- AWF: The AWF index will be calculated.

For example, if only PC and BCPC are to be calculated, and the input connection file is a *probability file*, the following needs to be written:

`-PC -BCPC`

The following would be an example of a case in which PC and F are to be calculated, and the input connection file is a *distance file*:

`-confProb 10000 0.5 -PC -F`

onlyoverall

This option specifies that only the overall index values for the entire landscape will be calculated (for all indices that have been selected in the previous option), and not the importance of each patch in maintaining connectivity of the network (percentage change in the index value after eliminating each patch, e.g. dPC).

The runs that specify *onlyoverall* will be completed much more quickly because calculations for the entire network only (e.g. PC for the overall landscape) are much faster than the calculations for obtaining the importance values of each individual patch (e.g. dPC for each node).

Note that BC(PC) is only an index at the node level, not at the overall landscape level. Therefore, even if it is selected for calculation, it will not be reported in the overall index files (this index will only be reported in the node importance file).

[-double | -simple]

Specifies the precision with which the calculations will be performed, with two options:

-simple: Single precision (float) is used to calculate the indices. Do not use this option when you want to obtain the importance values (deltas and fractions thereof) for each of the nodes (patches). Single precision is to be used only if you are only interested in obtaining the 'overall index values'.

-double: Uses high precision (double) to calculate the indices. This is the default precision that is used in the calculations if none of the two options (simple or double) is specified in the command line. The increased accuracy requires a somehow greater use of RAM (which should not be a problem except in very large graphs) and a little more time to process (although the increase in processing time compared to the "simple" precision is usually very small). The simple precision can be enough when you want to only calculate the total value of PC or other indices ("overall index values"), whereas the double precision should be used whenever you want

to get the node or link importance values, including for example the three fractions of dPC (dPCintra, dPCflux, dPCconnector); the simple precision can lead to less accurate calculations for these indices and fractions. In case of doubts, definitely use double precision.

-wprobdir

Writes the file with the direct dispersal probabilities (p_{ij}) among each pair of nodes. Note that in the Conefor version for directed graphs, this file will have $n(n-1)$ lines (where n is the number of nodes in the network), no matter if the input connection file is a full (*all*) or a partial (*notall*) connection file. When the input connection file is a partial file, the file written with this option will have a zero value for all the missing pairs of nodes or directions in such partial file (since, as noted earlier, pairs of nodes/directions missing in the input connection file are assumed to have a zero probability of direct dispersal).

-wprobmax

Writes the file with the maximum product probabilities (p^{*ij}) among each pair of nodes. For this file to be generated, the PC index must be selected for calculation. As above, note that in the Conefor version for directed graphs, this file will have $n(n-1)$ lines, no matter if the input connection file is a full (*all*) or a partial (*notall*) connection file. Note that, when the input connection file is a partial file, the file written may have a maximum product probability value even for those pairs of nodes or directions that were missing in such input partial connection file. This is because the maximum product probabilities (p^{*ij}) account for both direct and indirect connections between nodes ($p^{*ij} \geq p_{ij}$). Even when a zero probability of direct dispersal (p_{ij}) between two nodes i and j is assumed when this pair/direction is not listed in the input connection file, these two nodes i and j may still be connected indirectly, by using other intermediate stepping-stone nodes in a path comprising multiple links and nodes, resulting in p^{*ij} being larger than p_{ij} .

-noout

The processing information is not shown on the screen but rather is saved in the file `execution_events.txt`. This will typically result in faster processing because it takes usually less time to write to the file than to the screen.

Summary files for the results of multiple executions of the directed version of Conefor command line

Overall index values for the entire landscape

When (i) the option `-*` has been specified in a given execution of Conefor command line and (ii) all the files required for executing the Conefor command line (`conefor.exe`, node files, and connection files) are all located in the same folder, the results of the overall index values will be written in a single file with name `results_all_overall_indices.txt`. This file contains the results of all the multiple Conefor executions performed in that folder. This file will have one line for each execution, and of each index selected for calculation within the same execution, and will have the following five columns:

- Name of the prefix, as determined by the `-*` option (see above).
- Distance value specified in the option `-confProb distance prob` that should be used when the input connection file is a distance file (`-t dist`). If the input connection file is a probability file (`-t prob`), therefore not specifying any `-confProb distance prob` in the command line, Conefor will write 0 in this column.
- Probability value specified in the option `-confProb distance prob` that should be used when the input connection file is a distance file (`-t dist`). If the input connection file is a probability file (`-t prob`), therefore not specifying any `-confProb distance prob` in the command line, Conefor will write 0 in this column.
- Name of the calculated index (the overall value of which will be presented in the next column).
- Value of the calculated index.

Note that, if the PC index has been selected for calculation, the results of the overall index file will also include the values of the three fractions of the PC index (intra, direct, step) as described in Appendix S5 in Saura et al. (2014) (see reference above). Note that these three fractions are for the overall PC value for the entire landscape/network, and are different from the node-level fractions of dPC (intra, flux, connector) that were already implemented in previous versions of Conefor and can be found for each node in the node importance file (only if the option `"onlyoverall"` is not selected). The last paragraph of Appendix S5 in Saura et al. (2014) provides additional information on the differences between the overall landscape and the node-level fractions of PC.

Additionally, in the same case in which the option `-*` has been specified, one more file containing the values of the index EC(PC) (only if PC has been selected for calculation) will be written, with name `results_all_EC(PC).txt`. This file in fact just contains a part of the information already included in the file `results_all_overall_indices.txt` for all selected indices. The results of EC(PC) are saved separately because this is index if the one most frequently used when the overall

connectivity in the entire landscape is to be assessed. As above, this file will have one line for each execution and index selected for calculation within the same execution. The file will have the following four columns:

- Name of the prefix as determined through option *-**.
- Distance value specified in the option *-confProb distance prob* that should be used when the input connection file is a distance file (*-t dist*). If the input connection file is a probability file (*-t prob*), therefore not specifying any *-confProb distance prob* in the command line, Conefor will write 0 in this column.
- Probability value specified in the option *-confProb distance prob* that should be used when the input connection file is a distance file (*-t dist*). If the input connection file is a probability file (*-t prob*), therefore not specifying any *-confProb distance prob* in the command line, Conefor will write 0 in this column.
- Value of the EC(PC) index.

Sum of the PC importance values (dPC) for each individual node (node importances)

The results for each individual node (node importances) as calculated by PC will not be saved in a single joint summary file for multiple executions. They will be saved in an individual node importances file for each execution, which will have a different name as long as the option *-** is being used; otherwise, they will be saved overwriting the node importance file from previous executions.

This Conefor command line version will write however a summary of the results for the nodes importance values only for dPC and its fractions (dPCintra, dPCflux, dPCconnector). This summary consists in the sum of the dPC values for all nodes in the graph, as well as the sum of their fractions (dPCintra, dPCflux, dPCconnector) for all nodes in the graph. These results will be automatically saved in a file with name "node_importances_sum_PC.txt". Obviously, this will only happen if the PC index has been selected for calculation.

The file "node_importances_sum_PC.txt" will have the following seven columns, and one line for each execution of the Conefor command line in which PC has been selected for calculation:

- Prefix: name of the prefix (as given through the use of the *-** option). If option *-** is not used, the results will also be included in the file but writing "NA" in this column.
- Distance value specified in the option *-confProb distance prob* that should be used when the input connection file is a distance file (*-t dist*). If the input connection file is a probability file (*-t prob*), therefore not specifying any *-confProb distance prob* in the command line, Conefor will write 0 in this column.
- Probability value specified in the option *-confProb distance prob* that should be used when the input connection file is a distance file (*-t dist*). If the input connection file is a probability file (*-t prob*), therefore not specifying any *-confProb distance prob* in the command line, Conefor will write 0 in this column.
- Sum_dPC: sum of the dPC values for all nodes.
- Sum_dPCintra: sum of the dPCintra values for all nodes.
- Sum_dPCflux: sum of the dPCflux values for all nodes.
- Sum_dPCconnector: sum of the dPCconnector values for all nodes.