

How to Visualize Colexification Networks in Cytoscape (How to Do X in Linguistics 14)

Annika Tjuka
Department of Linguistic and Cultural Evolution
Max Planck Institute for Evolutionary Anthropology

The ability to visualize data in an intelligible way is an important skill for scientists. In linguistics, especially in lexical semantics, data are often visualized using graphs, i.e., networks. For example, in the web app for the Database of Cross-Linguistic Colexifications (CLICS), we use networks to illustrate that a lexical form refers to two different concepts by connecting the concepts (i.e., nodes) with a line (i.e., edge). When identifying the colexifications between concepts across a large number of languages, the network grows and a tool to visualize multiple data points becomes necessary. Here, I present a tutorial for the first steps to visualize a colexification network with Cytoscape. The tutorial is intended for beginners who want to learn how the tool works and serves as a starting point for further skill development.

1 Introduction

Data visualization is becoming increasingly important in science. A good graphic can facilitate the understanding of research results and workflows. It is part of communicating your research clearly and is an integral part of every journal article. There are many books on data visualization (e.g., Cairo 2016) and many tools for visualizing different types of data. Here, I focus on one tool that is often used to visualize networks: Cytoscape (Shannon et al. 2009). Besides Cytoscape, there is a comparable tool with similar features: Gephi (Bastian et al. 2009). Both are open-source network visualization tools. As far as my impression goes, Gephi offers a slightly more user-friendly interface and some additional modules for statistical analysis, while the Cytoscape community is larger since it is the standard tool for biologists and social scientists. Another option for visualizing networks is the networkX package in Python

(Hagberg et al. 2008), which is also used in the creation of the CLICS database (Rzymiski et al. 2020), but requires additional programming skills.

Networks are used to visualize the connection between two data points. In lexical semantics, we use networks to illustrate the relationship between two concepts. In recent years, there have been a large number of studies on colexifications, a phenomenon that occurs when a lexical form is used for two concepts in at least two genealogically unrelated languages (François 2008). All of these studies use networks to visualize their results (e.g., Jackson et al. 2019; Di Nataleet al. 2021; List 2023). Another way to visualize lexical data is through semantic maps, a type of network that illustrates the relationship between elementary meanings (Rakhilina et al. 2022). For example, Georgakopoulos and colleagues use Gephi to create their semantic maps (Georgakopoulos & Polis 2022; Georgakopoulos et al. 2022). The following examples are taken from unpublished studies on the colexification of body and object concepts.

2 First Steps

You can download the Cytoscape software here: <https://cytoscape.org/download.html>.

After downloading Cytoscape for your operating system, you open it. There are already sample sessions which you can explore. If you want to create your network from a file, you can use the Import Network from File option:

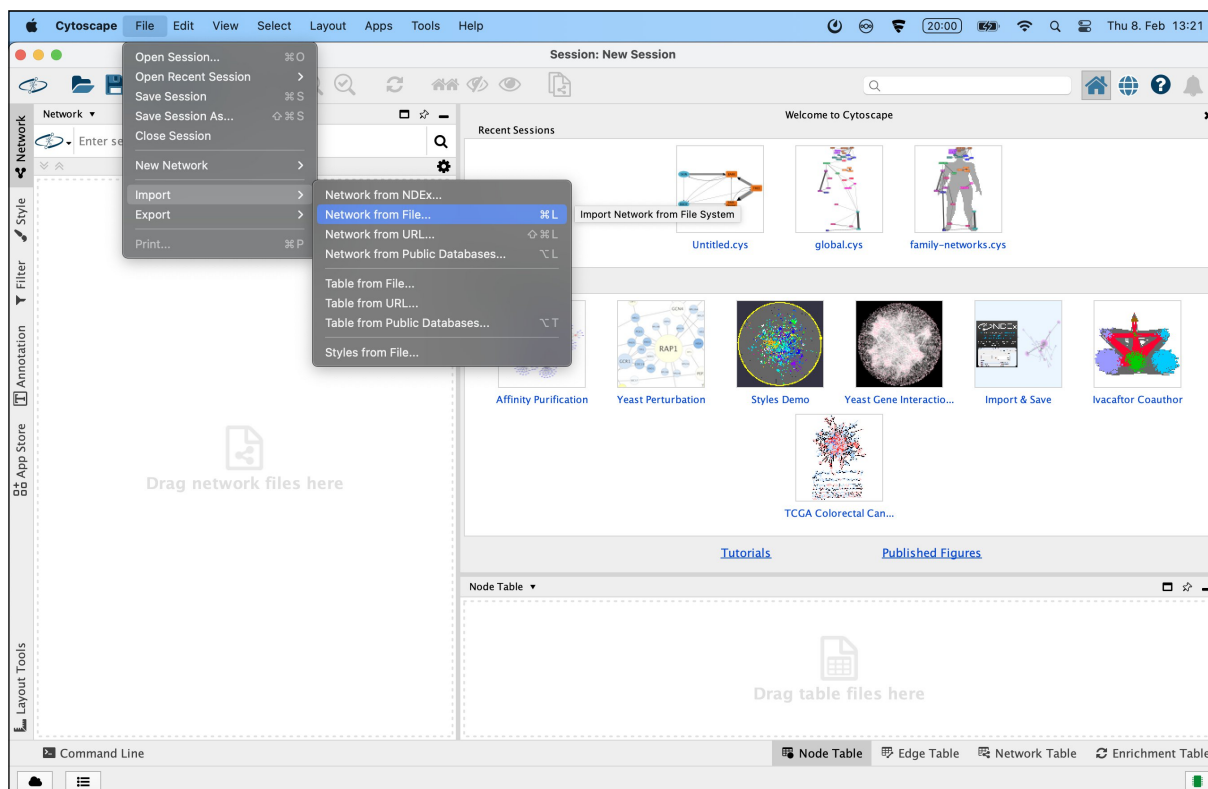


Figure 1. Importing a network into Cytoscape.

Possible file formats include .gml files or table formats such as .csv. When importing a .gml file the nodes and edges are already defined so the network is imported right away. If you use a simple table format, you will need to define the source and target node. For example, in the following list, Body would be the source (green dot) and Object the target (orange dot). The column Colexified includes the frequencies of colexifications between the two concepts across languages represented by the edges in a network.

Table 1. List of body-object colexifications across languages.

Body	Object	Colexified
SKIN	LEATHER	160
SKIN	BARK	90
TESTICLES	EGG	31
HEAD	TOP	24
NECK	COLLAR	24
BUTTOCKS	BOTTOM	15

The import looks like this:

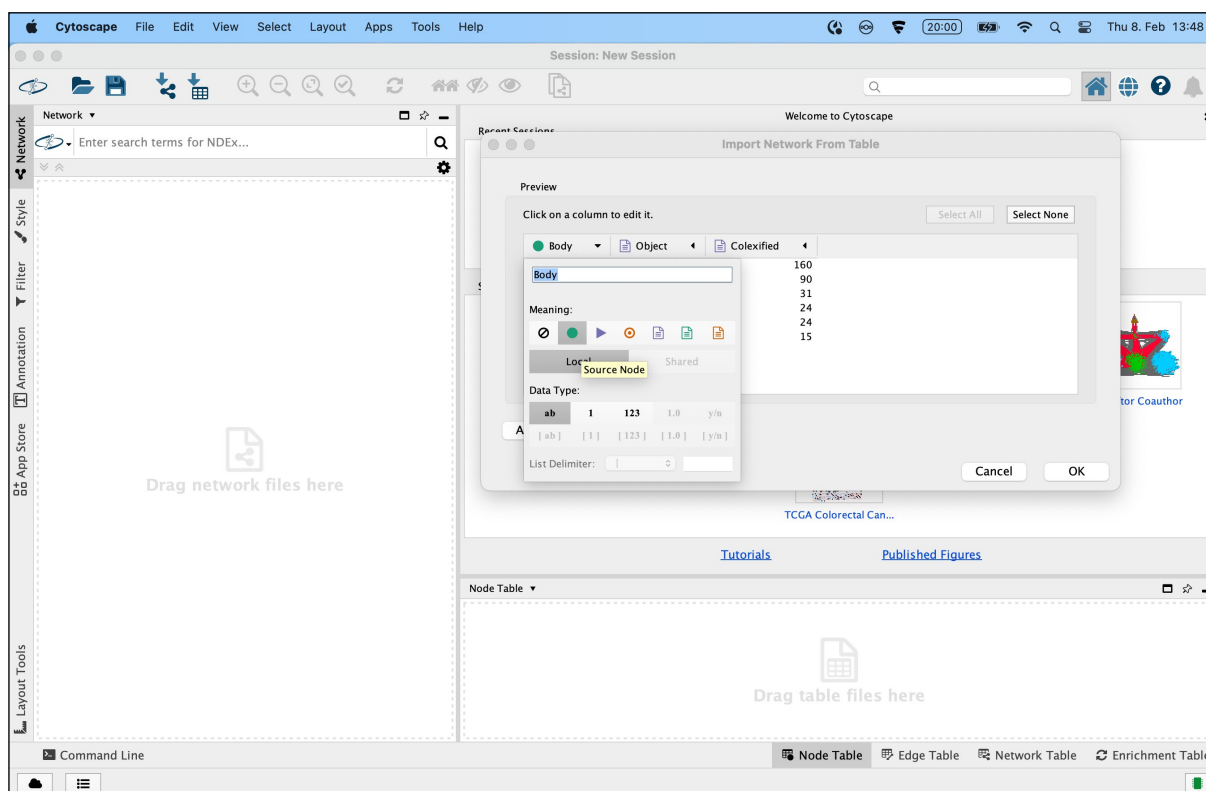


Figure 2. Define column for Source Node.

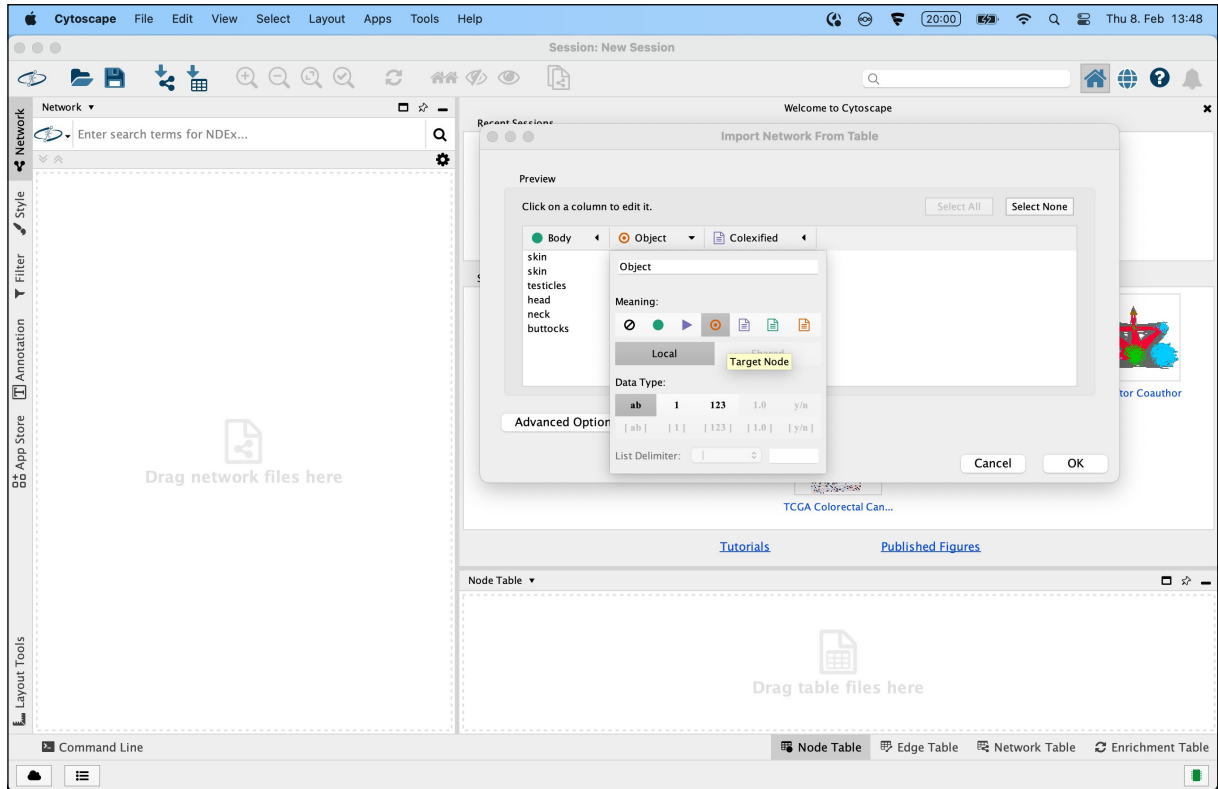


Figure 3. Define column for Target Node.

When pressing OK your first network appears on the right.

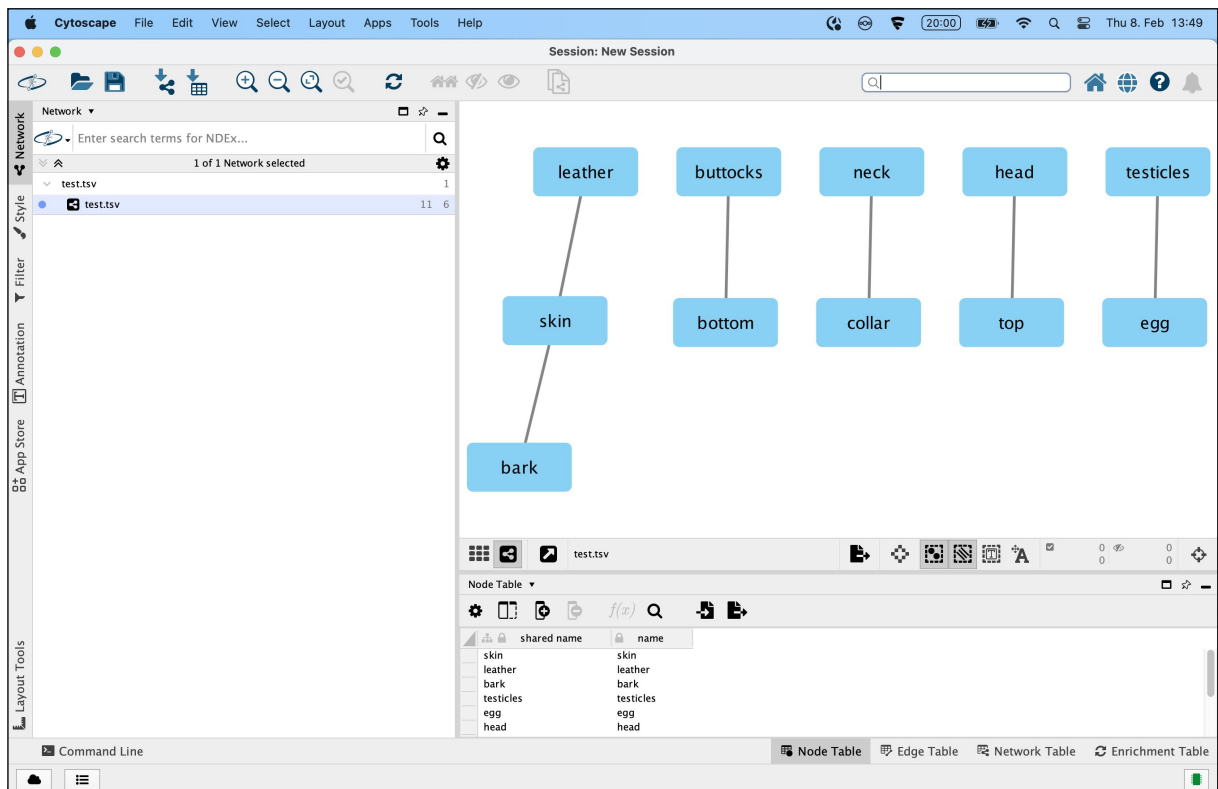


Figure 4. Simple collexification network without adjusting the style settings.

You can now explore the `Styles` tab where you can adjust the illustration of the network by either manually changing the style settings for the nodes and edges or choosing one of the predefined styles. If you make adjustments to the default style, you can save your personal style by selecting `Create New Style` under the tab `Options`. Below, I present some considerations when creating your network style.

3 Style Properties

When creating a network in Cytoscape, you should adhere to some general design principles:

- The labels need to be readable.
- The spatial arrangement should be meaningful. If possible, edges should not overlap.
- The choice of colors should be meaningful and not a random mix of bright colors. You can also consider a color-blind-friendly palette.
- It might be helpful to use different shapes to distinguish parts of your network.

An example of a network for colexifications between body and object concepts is shown below (Fig. 5). Note that I distinguished the body and object concepts by different colors and shapes. I also moved the colexifications with `LEAF` further away from the colexifications with `SKIN` since I discussed these in separate paragraphs in the article from which the network is taken (Tjuka 2024, forthcoming). You can move the network around in space by selecting one or multiple nodes.

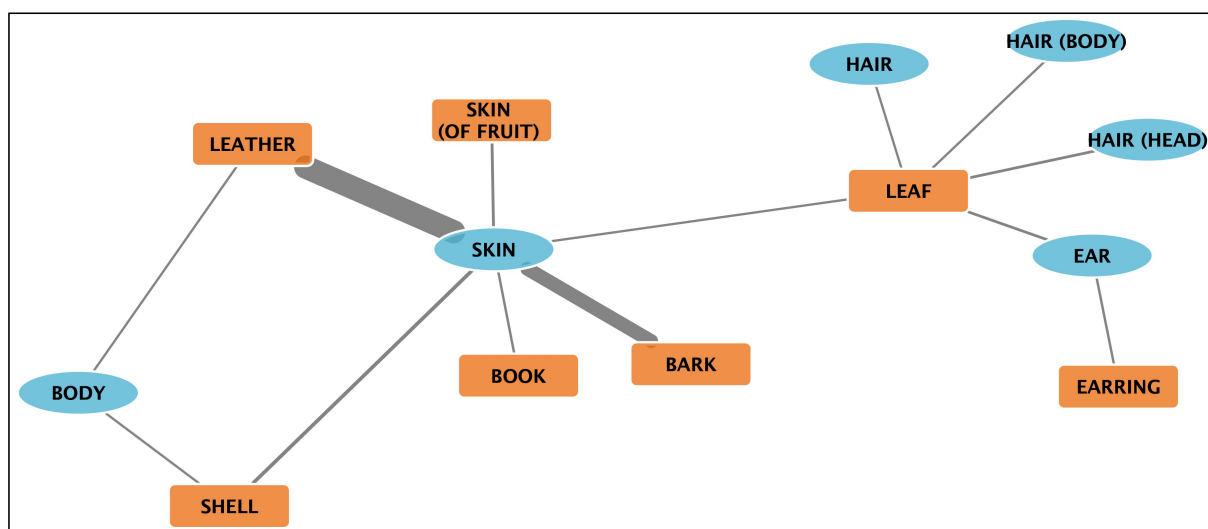


Figure 5. Example of a body-object colexification network.

Cytoscape gives you ample opportunity to modify the shape of your network. There are predefined properties that are shown on the left for the nodes and edges but you can also extend those with other properties:

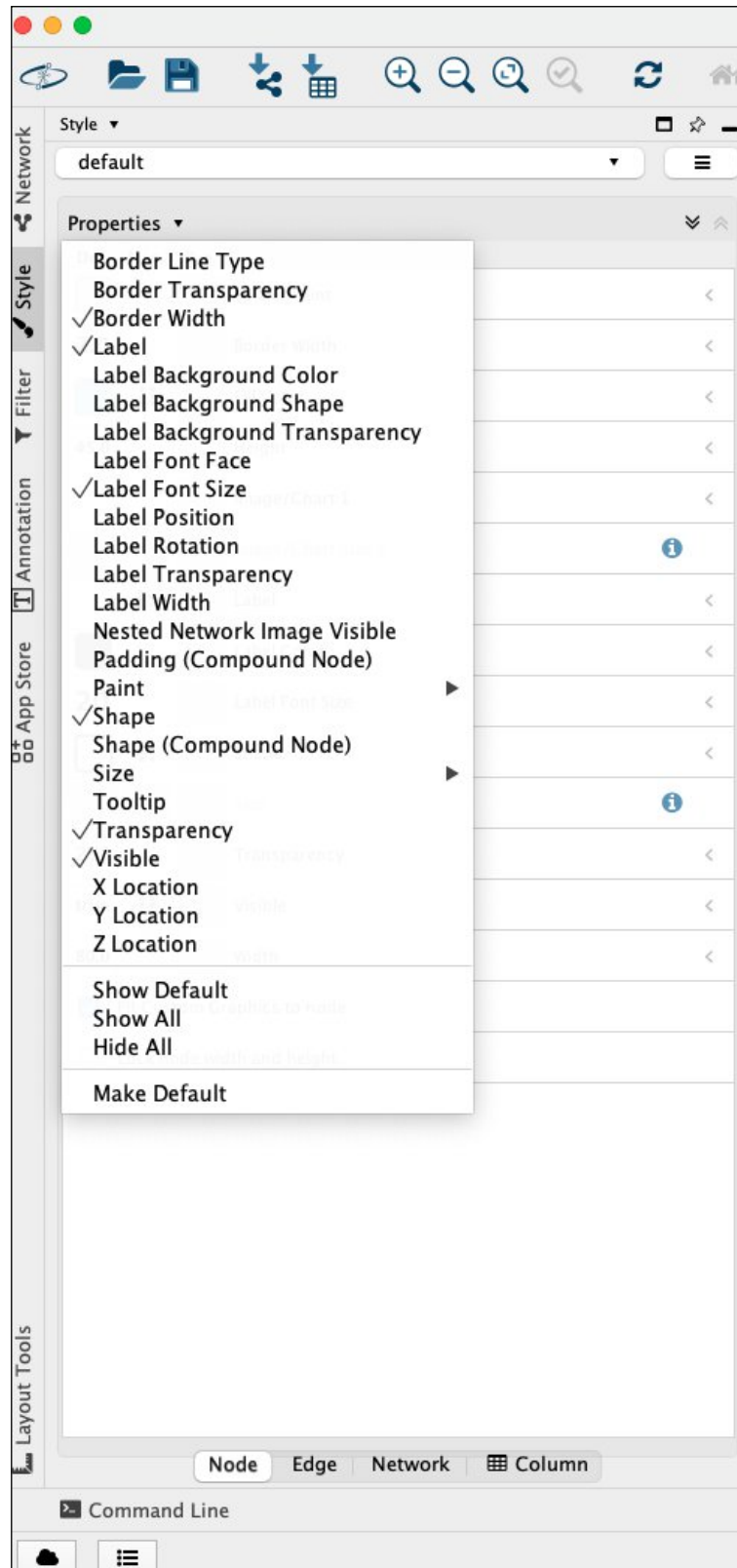


Figure 6. List of style properties available for nodes.

To create a weighted network such as the one above, an important feature is the `Width` property for the edges. Here, you can choose the frequency column, i.e., language count, the mapping type (`Continuous Mapping`), and the edge width. By double-clicking on the box with the triangle, a window opens where you can define the minimum and maximum.

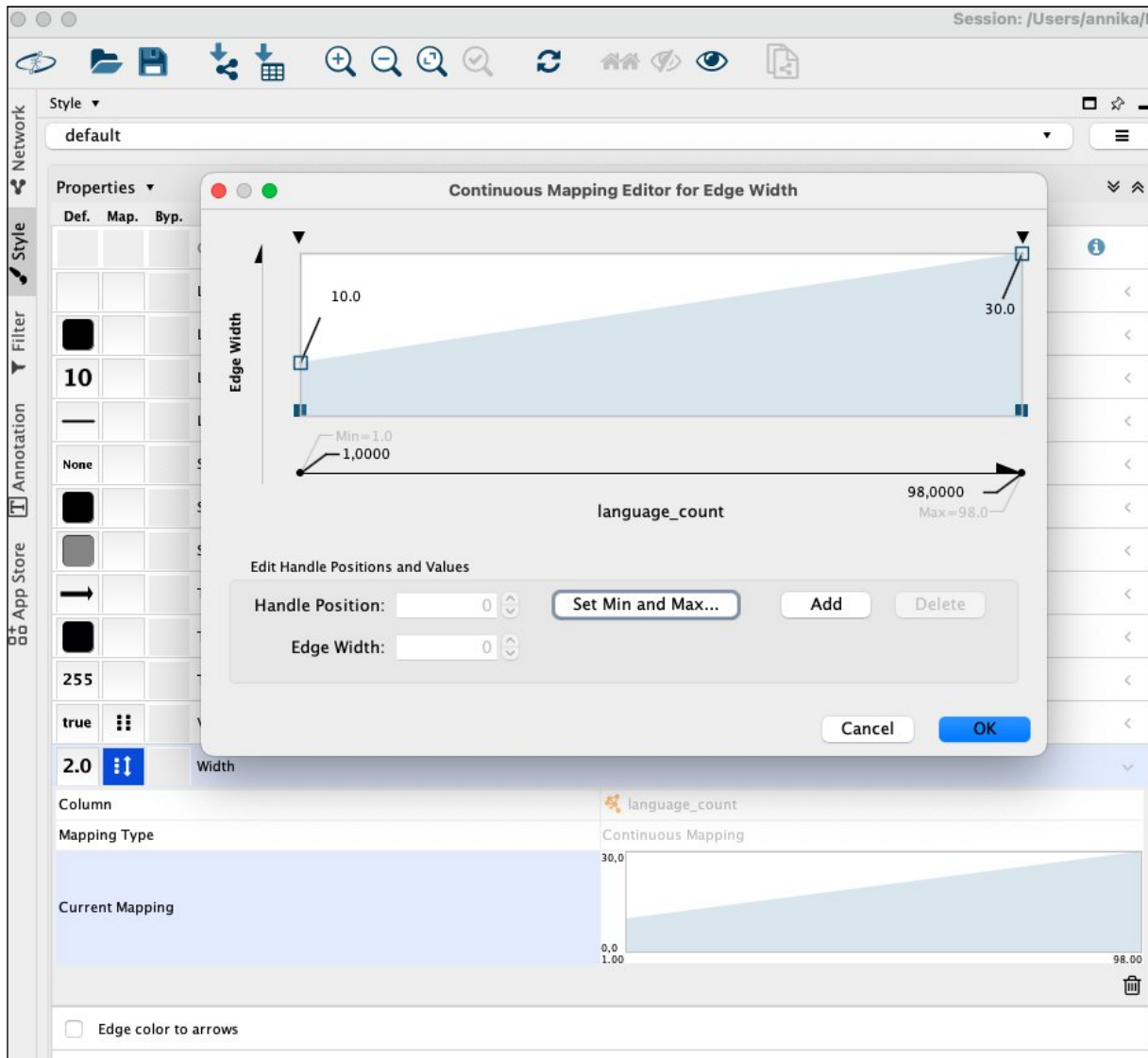


Figure 7. Adjusting the `Width` property when creating a weighted network.

With a directed network, you can apply the same principles but change the target arrow shape property. In directed networks, it is essential to arrange the nodes properly in space because they tend to become quite cluttered. An example is shown below (Fig. 8). Again, less is more, so you should be careful with colors and different arrow shapes. Some arrow shapes become less visible if the weight of the edge is low.

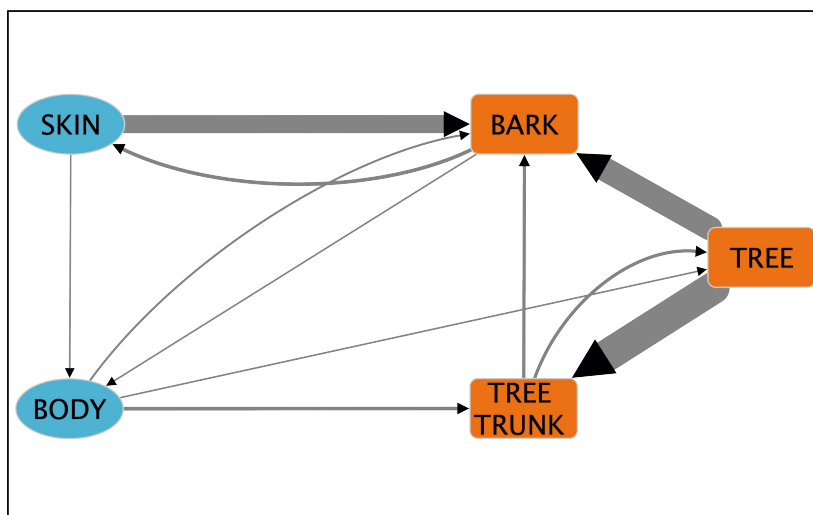


Figure 8. Example of a directed body-object colexification network.

The examples given in this section illustrate the possibilities for creating networks in Cytoscape. Depending on the network that you want to create, you can experiment with modifying the different properties. If you are happy with your network visualization, you can save the network as an image by clicking the file export icon on the right corner:

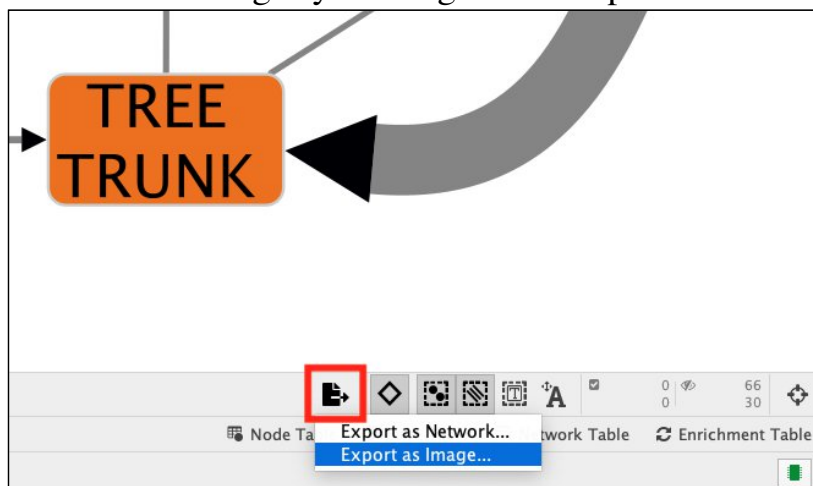


Figure 9. Exporting an image of the network.

You have multiple file options so if you want to further tweak your network, you can export it as a .svg file and import it, for example, into Inkscape.

4 Additional Resources

You can find more tips and tricks on how to create networks in Cytoscape on the internet. Here, I list the resources provided by the Cytoscape community that I found most useful.

- Manual: <https://manual.cytoscape.org/en/stable/>
- Tutorial: <https://github.com/cytoscape/cytoscape-tutorials/wiki>
- YouTube: <https://www.youtube.com/channel/UCv6auk9FK4NgXiXiqrDLccw>

5 Conclusion

Cytoscape offers many possibilities to visualize networks. One drawback of the many possibilities is that it is sometimes unclear where you can make which change. Or you fall down a rabbit hole where you change all the properties of the network with the result being an obscure visualization. It is therefore a good idea to always keep general design principles in mind and get feedback from colleagues as to whether your visualization is plausible for others.

References

- Bastian, Mathieu, Sebastien Heymann & Mathieu Jacomy. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. In *Proceedings of the International Conference on Weblogs and Social Media*. San Jose: Association for the Advancement of Artificial Intelligence. <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>.
- Cairo, Alberto. 2016. *The Truthful Art: Data, Charts, and Maps for Communication*. San Francisco: New Riders Press.
- Di Natale, Anna, Max Pellert & David Garcia. 2021. Colexification Networks Encode Affective Meaning. *Affective Science* 2. 99–111. <https://doi.org/10.1007/s42761-021-00033-1>.
- François, Alexandre. 2008. Semantic Maps and the Typology of Colexification: Intertwining Polysemous Networks Across Languages. In Martine Vanhove (ed.), *From Polysemy to Semantic Change: Towards a Typology of Lexical Semantic Associations*, 163–215. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.106.09fra>.
- Georgakopoulos, Thanasis, Eitan Grossman, Dmitry Nikolaev & Stéphane Polis. 2022. Universal and Macro-Areal Patterns in the Lexicon: A Case-Study in the Perception-Cognition Domain. *Linguistic Typology* 26(2). 439–487. <https://doi.org/10.1515/lingty-2021-2088>.
- Georgakopoulos, Thanasis & Stéphane Polis. 2022. New Avenues and Challenges in Semantic Map Research (with a Case Study in the Semantic Field of Emotions). *Zeitschrift für Sprachwissenschaft (Special Issue on the Future of Mapping: New Avenues for Semantic Maps Research)* 41(1). 1–30. <https://doi.org/10.1515/zfs-2021-2039>.
- Hagberg, Aric A., Daniel A. Schult & Pieter J. Swart. 2008. Exploring Network Structure, Dynamics, and Function Using NetworkX. In Gael Varoquaux, Travis Vaught & Jarrod Millman (eds.), *Proceedings of the 7th Python in Science Conference*, 11–15. Pasadena.
- Jackson, Joshua Conrad, Joseph Watts, Teague R. Henry, Johann-Mattis List, Robert Forkel, Peter J. Mucha, Simon J. Greenhill, Russell D. Gray & Kristen A. Lindquist. 2019. Emotion Semantics Show Both Cultural Variation and Universal Structure. *Science* 366. 1517–1522. <https://doi.org/10.1126/science.aaw8160>.
- List, Johann-Mattis. 2023. Inference of Partial Colexifications from Multilingual Wordlists. *Frontiers in Psychology* 14. 1–10. <https://doi.org/10.3389/fpsyg.2023.1156540>.
- Rakhilina, Ekaterina, Daria Ryzhova & Yulia Badryzlova. 2022. Lexical Typology and Semantic Maps: Perspectives and Challenges. *Zeitschrift für Sprachwissenschaft (Special Issue on the Future of Mapping: New Avenues for Semantic Maps Research)* 41(1). 231–262. <https://doi.org/10.1515/zfs-2021-2046>.
- Rzyski, Christoph, Tiago Tresoldi, Simon J. Greenhill, Mei-Shin Wu, Nathanael Schweikhard, Maria Koptjevskaja-Tamm, Volker Gast, et al. 2020. The Database of Cross-Linguistic Colexifications, Reproducible Analysis of Cross-Linguistic Polysemies. *Scientific Data* 7(1). 1–12. <https://doi.org/10.1038/s41597-019-0341-x>.
- Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski & Trey Ideker. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* 13(11). 2498–2504. <https://doi.org/10.1101/gr.1239303>.

Tjuka, Annika. (2024, forthcoming). Objects as Human Bodies: Cross-Linguistic Colexifications Between Words for Body Parts and Objects.

Supplementary Material

The list of colexifications and an example network for further exploration is available on Zenodo: <https://doi.org/10.5281/zenodo.10649258>

Funding Information

This project has received funding from the Max Planck Society as part of the CALC³ Research Project (<https://calc.digling.org>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.