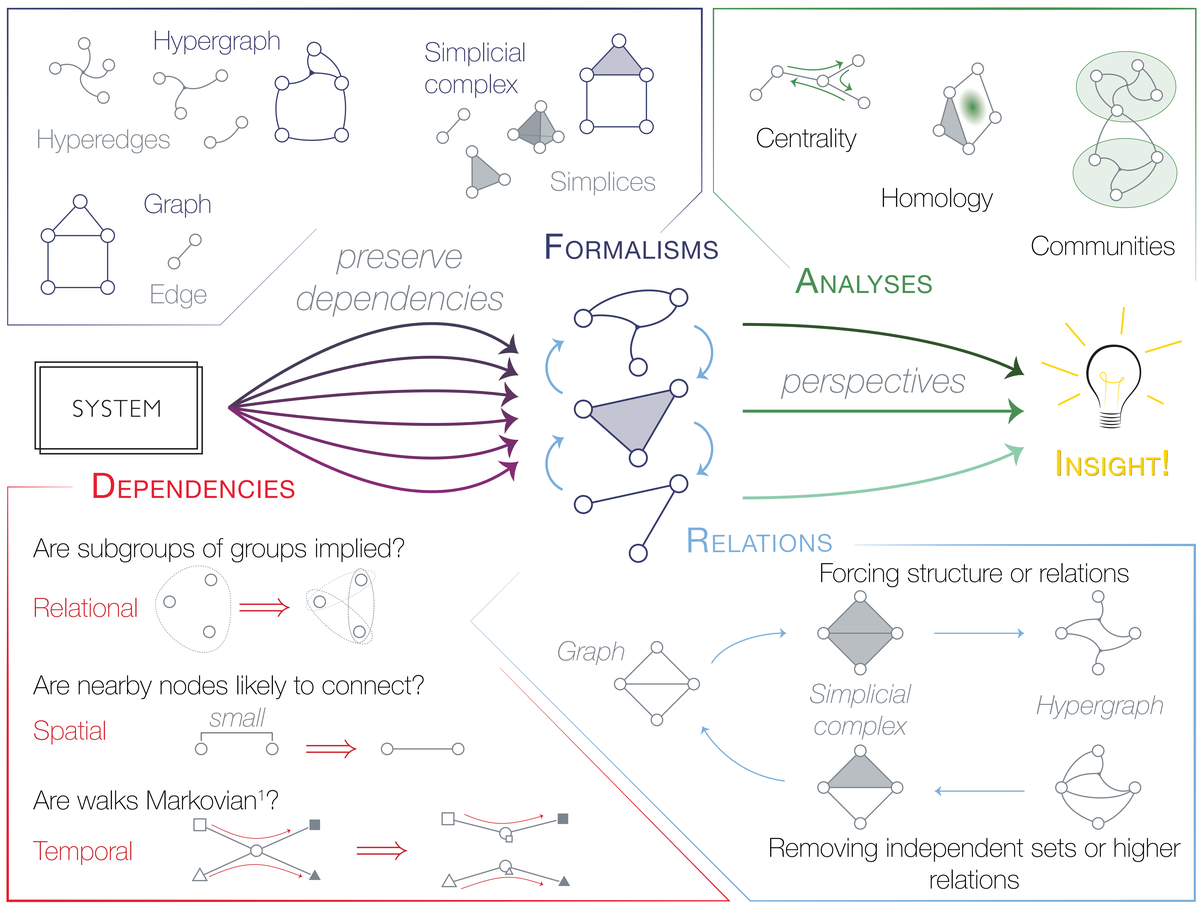
The why, how, and when of representations for complex systems

Leo Torres[1] and Ann Sizemore Blevins[2], Danielle S. Bassett[2], Tina Eliassi-Rad[1]
[1]Northeastern University, [2]University of Pennsylvania

How do we analyze a complex system?

Complex systems abound in many disciplines from neuroscience and computer science to ecology and economics. Despite (or perhaps due to) their prevalence, researchers often have different answers for how best to represent, encode, and analyze a complex system. Here we colelct analysis frameworks, highlight assumptions made within pipelines, and distinguish use cases.



Formalisms offer different perspectives

Hyperedges can have absent substructures. We measure the fullness of a hyperedge with the fill coefficient:

$$f(h) = \frac{|g \in E : g \subsetneq h \text{ and } |g| > 1|}{2^{|h|} - 2 - |h|}$$

Similar to the graph formalism, we can calculate the hypergraph clustering coefficient[2]:

$$HC(v) \propto \sum_{h_i, h_j, v \in h_*} EO(h_i, h_j)$$

$EO(h, g) =$ extra overlap between hyperedges

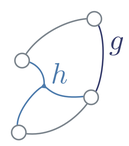
Does our interpretation of average clustering change when using different representations?

Hypergraph

$EO(h, g) = 2/3$

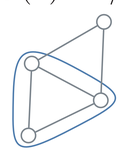
$f(h) = 1/3$

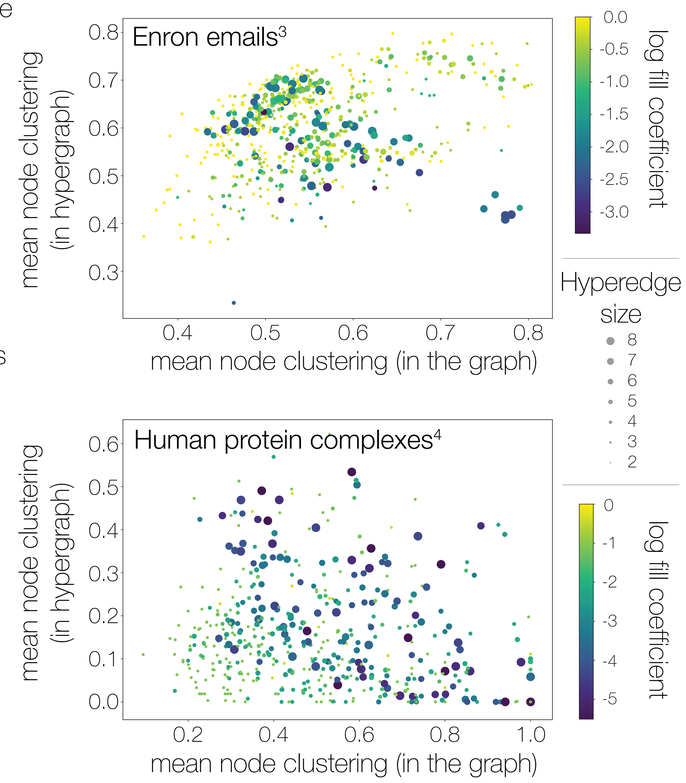
$\overline{HC}(h) = 8/27$



Graph

$\overline{C}(h) = 7/9$





Conclusions

- Choosing the proper formalism for complex system analyses requires knowledge of data dependencies and question requirements.
- The data abstraction method can significantly affect analysis results.

References

1. Benson, Austin R., David F. Gleich, and Jure Leskovec. "Higher-order organization of complex networks." Science 353.6295 (2016): 163-166.
2. Zhou, Wanding, and Luay Nakhleh. "Properties of metabolic graphs: biological organization or representation artifacts?." BMC bioinformatics 12.1 (2011): 132.
3. Klimt, Bryan, and Yiming Yang. "The enron corpus: A new dataset for email classification research." European Conference on Machine Learning. Springer, Berlin, Heidelberg, 2004.
4. Giurgiu, Madalina, et al. "CORUM: the comprehensive resource of mammalian protein complexes—2019." Nucleic acids research 47.D1 (2018): D559-D563.