Mapping flows on hypergraphs: How choosing random-walk model and network representation matters for community detection

Anton Eriksson,[1,*] Daniel Edler,[1] Alexis Rojas,[1] Manlio de Domenico,[2] and Martin Rosvall[1]
[1] *Integrated Science Lab,*
Department of Physics,
Umeå University, SE-901 87 Umeå,
Sweden
[2] *CoMuNe Lab, Fondazione Bruno Kessler,*
Via Sommarive 18, 38123 Povo (TN),
Italy

Hypergraphs offer an explicit formalism to describe multi-body interactions in complex systems. Network scientists have generalized random-walk models to hypergraphs to connect dynamics and function in systems with these higher-order interactions and studied the multi-body effects on flow-based centrality measures. But mapping the large-scale structure of those flows requires effective community detection methods.

While several methods can identify flow-based communities in multilayer and memory networks with non-Markovian dynamics, researchers have just begun to unravel the large-scale systemic effects of multi-body interactions captured by hypergraphs. However, different systems and research questions call for different random walk and hypergraph models: Random walks can be lazy, able to visit the same node multiple times in a row, or non-lazy and forced to move on. Hyperedges can have arbitrary weights, and nodes can have hyperedge-dependent weights. Because these and other models can be represented with different network types – bipartite, unipartite, and multilayer – the questions multiply: For different data and different questions, which model and representation is best?

To address which combination of model and representation is best for answering different questions about various hypergraph data, we derive unipartite, bipartite, and multilayer network representations of hypergraph flows with identical node-visit rates for the same random-walk model. For unique node-visit rates when a representation requires directed links, we apply an unrecorded teleportation scheme robust to changes in the teleportation rate and preserve the node-visit rates when teleportation is superfluous in undirected networks. The information-theoretic and flow-based community detection method Infomap allows us to explore how different hypergraph random-walk models and network representation change the number, size, depth, and overlap of identified multilevel communities.

By analyzing schematic and real hypergraphs, we find that the bipartite network representation requires the fewest links and enables the fastest community detection. A multilayer network representation that reinforces flows within similar layers gives the deepest modular structures with the most overlapping communities but at a high computational cost. The unipartite network representation provides a trade-off between the two, with intermediate compactness, speed, and detectable modular regularities.

We compare our representations by mapping network science researchers into research groups grouped into research areas. Finally, we create a map of marine fossil occurrences from the Cambrian to the Cretaceous period to highlight macroevolutionary transitions.

* anton.eriksson@umu.se

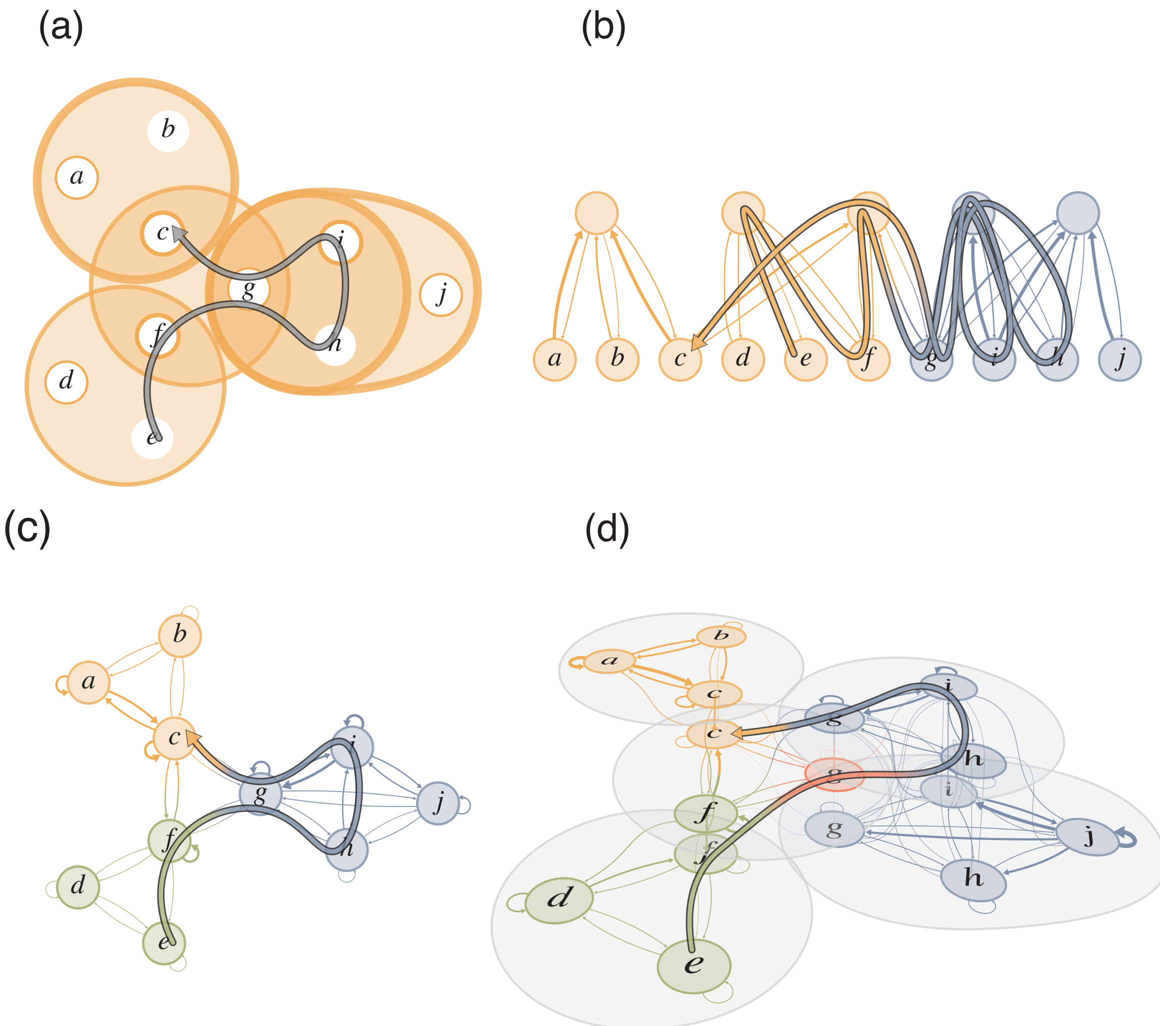

Fig. 1. A schematic hypergraph represented with three types of networks. (a) The schematic hypergraph with weighted hyperedges and hyperedge-dependent node weights. Thin borders for weight 1 and thick borders for weight 3. A lazy random walk on the schematic hypergraph represented on: (b) a bipartite network, (c) a unipartite network, and (d) a multilevel network. The colors indicate optimized module assignments, in (d) for hyperedge-similarity walks.