Characterizing cellular response is a central question in systems biology, and signaling pathways describe this organization of molecular interactions. Databases such as NetPath, KEGG, and NCI-PID store well-studied pathways that are represented as networks. Given that these pathways serve to transmit extrinsic signals to alter gene expression, signaling pathways may show similar structure that reflects this common function. Further, signaling pathway topology may be different from networks consisting of more general interactions (e.g., from a background protein-protein interactome). We pose two questions:

**Question 1:** Are pathways topologically similar?

**Question 2:** Can we distinguish pathways from an interactome?

### III. Finding Over/Under-Represented Graphlets

We used multiple realizations of two distinct null model networks to assess the topological similarity of pathways.

(i) Random rewiring: Pairs of edges are randomly rewired, preserving the node degree of the pathway.

(ii) Random-Walker (RW)-Induced: A subgraph of the interactome constructed by taking the induced graph over the nodes visited by a random walker, preserving the number of nodes and edges in the pathway.

The figure below shows graphlet count in the NetPath EGFR1 pathway (red circles) compared with 100 realizations of the two different null models (boxplots).

We consider the graphlet count to be statistically over-represented (or under-represented) if it is at least two standard deviations above (or below) the respective mean count of the null model. We count graphlets for all pathways in each database, recording the fraction of pathways in which each graphlet appears either over-represented or under-represented.

### IV. Distinguishing Pathways from the Interactome

For each database, we generated a set of RW-Induced graphs with the same number of nodes and edges as its constitutive pathways. We then clustered the pathways and the RW-Induced networks to see if pathways clustered together. We show the dendrogram for agglomerative clustering with cosine similarity for NetPath below.

### V. Conclusions & Future Directions

Graphlets can quantify topological similarities of signaling pathways and distinguish them from background interaction networks.

1. Graphlet over/underrepresentation is not conserved across databases. Can graphlets characterize database-specific topologies?

2. Our methodology is intentionally simple. Can directed graphlets, more sophisticated similarity measures, or more complex clustering algorithms better identify pathway topology?

3. Graphlets show promise for discovering under-studied pathways within interactomes. Can we find subgraphs of the interactome that reflect pathway-specific graphlet distributions?

Acknowledgements: We thank Anthony Gitter and Chris Magnano for discussions about this research. This work was supported by the National Science Foundation (BIO-ABI #1750981 to Anna Ritz).

---

**References**


