Finding Higher Structures in Graphs

Abdullah Malik, Major Professors: Dr. Tyler Foster, Dr. Mark van Hoeij

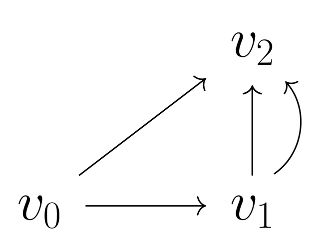
Pure Mathematics, Department of Mathematics, Florida State University

Abstract

Real world graphs model bidirectional relationships, often destroying information about any higher relations. These are rather modelled using hypergraphs. In [1], hypergraphs are formulated using the language of simplicial sets. Here, we propose a technique that recovers the higher structure within a graph by formulating a graph as simplicial set. We use the adjacency matrix of the graph to discover these higher structures. This approach has the advantage of potentially speeding up walk over the graph, since GPUs are designed to handle matrix computations efficiently, promising use for machine learning algorithms [2].

Motivation

In a directed citation graph where nodes are authors, and a directed edge from vertex a to vertex b represents "author a cites author b" in one paper, it is unclear if the graph



represents a) two coauthors v_0, v_1 citing papers authored by v_1 and v_2 individually, or b) if there are four papers with single authors each, with v_2 cited in both papers authored by v_1.

Definitions and Formulation

We define a **multidirected graph** as a functor $G^{op} \in Set^{[0] \rightrightarrows [1]}$ from the Walking arrow category $\{[0] \rightrightarrows [1]\}$ to the category $\mathcal{S}et$, where $G(0) = V$ and $G(1) = E$ with source and target maps $s, t : E \longrightarrow V$. Edges with multiplicity n are labelled by $e^{(k)} = [v_i, v_j]^{(k)}$ for $k \in [n] = \{0, 1, ..., n\}$ with $s\left(e^{(k)}\right) = v_i$ and $t\left(e^{(k)}\right) = v_j$ for all k.

The Walking arrow category $\Delta_{<2}$ is a subcategory of Δ, the (skeleton) category with finite sets $[n]$ as objects and monotone maps as morphisms.

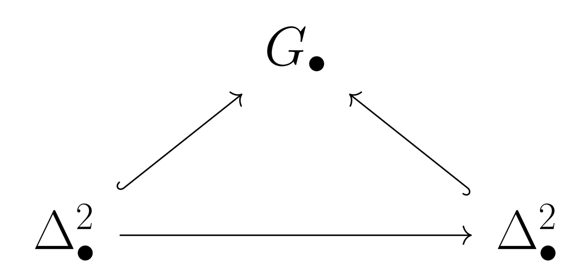
A functor $X^{op} \in Set^{\Delta}$ is called a simplicial set. The inclusion $i : \Delta_{<2} \hookrightarrow \Delta$ induces a functor $R_i : \mathcal{S}Sets = Sets^{\Delta} \longrightarrow Sets^{\Delta_{<2}} = \mathcal{G}$ which sends simplicial sets X^{op} to their 1-skeleton $R_i(X) = X^{op} \circ i$.

There is also another functor $L_i : \mathcal{G} \longrightarrow \mathcal{S}Sets$ and a natural bijection $Hom_{\mathcal{G}}(G, R_i(X^{op})) \cong Hom_{\mathcal{S}Set}(L_i(G), X^{op})$. Here, L_i is the left Kan extension of X^{op} along i, so $L_i(G)([n]) = \text{colim}_{[n] \longrightarrow i(x)} G(x)$. Concretely, getting a higher order structure implicit in a graph depends on how we visualise the collapse $[n] \longrightarrow i(x)$.

Let $L_i(\mathcal{G}) = G_\bullet$, where $G_\bullet([0]) = G_0 = \{[v_i]\}$ is the set of vertices, whereas $G_1 = \left\{[v_i, v_j]^{(x)}\right\}$ is the set of edges $[v_i, v_j]^{(x)} = e^{(x)}$. For $k \geq 1$, we also have $G_k = \left\{[v_{i_1}, v_{i_2}, ..., v_{i_k}]^{(x_k)}\right\}$ with $d_j\left([v_{i_1}, v_{i_2}, ..., v_{i_k}]^{(x_k)}\right) = \left[v_{i_1}, v_{i_2}, ..., \widehat{v_{i_j}}, ..., v_{i_k}\right]^{(x_k - 1)}$, for $j \leq k$. The choice of the collapse is reflected in the choice of x_k.

Example

For the graph on the left, we have 0-simplices $\{v_0, v_1, v_2\}$, nondegenerate 1-simplices $\left\{[v_0, v_1]^{(1)}, [v_0, v_2]^{(1)}, [v_1, v_2]^{(1)}, [v_1, v_2]^{(2)}\right\}$ and non-degenerate 2-simplices $\left\{[v_0, v_1, v_2]^{(1)}, [v_0, v_1, v_2]^{(2)}\right\}$, from the colimit



where $\Delta^2_\bullet$ is the standard 2-simplex and $\Delta^2_\bullet \xrightarrow{f_\bullet} \Delta^2_\bullet$ is defined by $f_0 = id_{X_0}$, and f_1 identity on $[0, 1]$ and $[0, 2]$, only. The geometric realization of this simplicial set is a cone. Therefore, there are two papers authored by v_1, both of which cite v_2 (the situation (b) on the left).

Finding the higher structure

Let A be the adjacency matrix associated with the multidirected graph $G_\bullet$ with $|G_0| = g$ vertices. By that, we mean that the entry $a_{ij}^{(1)} = n$ if there are n edges from vertex v_i to vertex v_j, and zero otherwise.

1. **Base Case** $(k = 2)$ For a given i, j, if $a_{ij}^{(2)} = \sum_{k=1}^{g} a_{ik}^{(1)} a_{kj}^{(1)} = \ell \neq 0$ in A^2, then $\exists k_1, k_2 ..., k_\ell \in \mathbb{N}$ such that $a_{ik_\alpha}^{(1)} = a_{k_\alpha j}^{(1)} = 1$, where $1 \leq \alpha \leq \ell$, and $1 \leq k_\alpha \leq g$. If $[v_i, v_j] \in G_1$, then $[v_i, v_{k_\alpha}, v_j] \in G_2$ for all α.
2. **Base Case** $(k = 3)$ A nonzero $a_{ij}^{(3)} = \sum_{\alpha=1}^{g} a_{i\alpha}^{(1)} a_{\alpha j}^{(2)} = \sum_{\beta=1}^{g} a_{i\beta}^{(2)} a_{\beta j}^{(1)}$ tells us that $\exists \alpha, \beta \in \mathbb{N}$ such that $a_{i\alpha}^{(1)}$ and $a_{\alpha j}^{(2)}$ are nonzero, or that $a_{i\beta}^{(2)}$ and $a_{\beta j}^{(1)}$ are nonzero. To determine α, we look for common indices in row i of A and column j of A^2. A sufficient but not necessary condition for the existence of a 3-simplex is the existence of at least 1 α. This can be used as a check to truncate unnecessary computations. To determine β, we similarly look for common indices in row i of A^2 and column j of A. Then $[v_i, v_\alpha, v_\beta, v_j] \in G_3$
3. **Inductive Step** Assuming that we have found $(n-1)$-simplices, and if $a_{ij}^{(n)} = \sum_{\alpha=1}^{g} a_{i\alpha}^{(1)} a_{\alpha j}^{(n-1)} = \sum_{\beta=1}^{g} a_{i\beta}^{(n-1)} a_{\beta j}^{(1)} = \ell > 0$, then we find α by looking at common indices of row i of A and column j of A^{n-1}. With β determined similarly, we need to find x_i's such that $d_k[v_i, v_\alpha, x_1, x_2, ..., x_{n-3}, v_\beta, v_j] \in G_{n-1}$ for all k. This search list comes from entries in G_{n-1} that begin with v_i or v_α, and end with v_β or v_j. In addition, all of these entries must be in row i of A (or column j of A). From this shortened list, we simply search for entries corresponding to $[x_1, x_2, ..., x_{n-3}] \in G_{n-4}$. All such entries will give us $[v_i, v_\alpha, x_1, x_2, ..., x_{n-3}, v_\beta, v_j] \in G_n$

Psuedocode

We specify maximum dimension d of simplices to look for. The algorithm terminates if there are no k-simplices for $k < d$. B is a list of matrices. The dictionary `skeleton` is keyed with dimensions and valued as vertices in list form. A is a coordinate matrix.

```
i_rows, i_columns ← non-zero indices of B[0]
for k from 2 to d do
    B[k] ← A^k
    n_rows, n_columns ← non-zero indices of B[k]
    p_rows, p_columns ← non-zero indices of B[k − 1]
    for i, j from n_rows, n_columns do
        o_smplces, i_smplces, o_nghbrs, i_nghbrs ← ∅
        for k, l from p_rows, p_columns do
            if i = k then o_smplces ← o_smplces ∪{l}
            if j = l then i_smplces ← i_smplces ∪{k}
        for k, l from i_rows, i_columns do
            if i = k then o_nghbrs ← o_nghbrs ∪{l}
            if j = l then i_nghbrs ← i_nghbrs ∪{k}
        B_1 ← o_smplces ∩ i_nghbrs, B_2 ← i_smplces ∩ o_nghbrs
        if B_1 = ∅ or B_2 = ∅ then return highest-dimension ← k
        if k>3 then
            I ← ∅                                          ▷ I = indices
            for v in skeleton[k-1] do
                if v[k]=j & v[k-1] ∈ B_2 & (v[0] = i ∨ v[0] ∈ B_1) then
                    if v[0] ∈ B_1 then α ←v[0], I ← I ∪{v}
                if v[0]=i & v[1] ∈ B_1 & (v[k-1] = j ∨ v[k-1] ∈ B_2)
    then
                    if v[k-1] ∈ B_2 then β ←v[k-1], I ← I ∪{v}
            I ← o_nghbrs ∩ I
            for x_1,x_2,...,x_{k−3} in I × I × ... × I do
                if x_1, x_2, ..., x_{k−3} in skeleton[k − 3] then skeleton ←
    skeleton+{k : i, α, x_1, x_2, ..., x_{k−3}, β, j}
```

References

[1] Spivak, D.I., 2009. Higher-dimensional models of networks. arXiv preprint arXiv:0909.4314.

[2] Bodnar, C., Frasca, F., Otter, N., Wang, Y., Lio, P., Montufar, G.F. and Bronstein, M., 2021. Weisfeiler and lehman go cellular: Cw networks. Advances in Neural Information Processing Systems, 34, pp.2625-2640.