Introduction

Graphlets are small connected induced subgraphs of a larger graph. Typically graphlets are said to be subgraphs containing up to 5 nodes, but they can be easily expanded to contain 6 or more nodes. Counting how many times each orbit (i.e. each symmetrically equivalent set of nodes) of each graphlet touches each node of a large graph is a good way to identify the type of network the graph represents.

Orbits can be counted more efficiently using equations, reducing counting orbits on \(n\) nodes to searching graphlets on \(n-1\) nodes and solving a set of equations involving sets of common neighbors of the nodes of these graphlets. The equations reflect the ways a graphlet can be built. Therefore, one orbit may have multiple equations. We investigated the effect of choice of equations on the speed of our Jesse algorithm.

Effect of equation choice

<table>
<thead>
<tr>
<th>Graphlet order</th>
<th>Counting time (s)</th>
<th>Speedup factor (low degree)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Random equation</td>
<td>Low degree added node</td>
</tr>
<tr>
<td>5</td>
<td>0.76</td>
<td>0.41</td>
</tr>
<tr>
<td>6</td>
<td>2196</td>
<td>9.85</td>
</tr>
<tr>
<td>7</td>
<td>826.25</td>
<td>349.00</td>
</tr>
</tbody>
</table>

Running time of the Jesse algorithm with different selection of equations. Whenever multiple equations can be used to count one orbit, the used equation is chosen at random, the equation in which the added node has the lowest degree, or the highest degree.

But why?

**Lookup in hash maps with high collision rate?**

No, because the speed difference still exists when all common neighbors are saved in the same hash map.

**Comparison of longer lists?**

No, because determining equality of lists only based on hash (only computed once) does not remove the speed difference.

**Just more lookups?**

Appears to be the case, but the time is not directly proportional to the number of lookups. Maybe because adding the least connected node last assures that the starting graphlet is the most dense possible, therefore rarer? As is to be expected in that case, the speed difference vanishes in extremely dense and sparse graphs.

Further reading

Ine Melckenbeeck, Pieter Audenaert, Tom Michael, Dieter Colle, Mario Pickavet (2016); An Algorithm to Automatically Generate the Combinatorial Orbit Counting Equations. PLOS ONE 11(1): e0147078.

Ine Melckenbeeck, Pieter Audenaert, Didier Colle, Mario Pickavet (2017); Efficiently counting all orbits of graphlets of any order in a graph using autogenerated equations, Bioinformatics (accepted, to be published soon): btx758