7 Machine Learning and Real-world Data (sht25)

Consider the directed graph shown in the figure below, which expresses cooperation amongst individuals (A, B, . . . , H) in a fishing village. The meaning of an edge from X to Y is that X has asked Y for advice or help during fishing at least once.

(a) Consider the betweenness centrality of each individual in this network, which is listed in the following table.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>12</td>
<td>2</td>
<td>9</td>
<td>0</td>
<td>12</td>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

(i) Give a definition of the betweenness centrality of a node. [1 mark]

(ii) Explain intuitively why B and F have the highest betweenness centralities and why E and H have betweenness centralities of 0. [2 marks]

(b) We now look at what happens if the network is converted into an undirected network.

(i) What is the diameter of this network and why? Your question should include a definition of diameter. [2 marks]

(ii) Do the betweenness centralities of nodes A and C change, and why? Explain in terms of affected paths. [3 marks]

(iii) Consider the general case of two near-identical graphs S and T, where S is a directed graph and T is the undirected version of S, i.e., every edge \((u, v)\) in S is replaced by an undirected edge \((u, v)\) in T. Which of the following statements are true about the betweenness centrality of any pair of nodes \(X_S\) and \(X_T\), which are in identical relative position in the graphs? Justify your answer or provide a counter example.

[continued ...]
(A) The betweenness centrality of $X_S$ is always at least that of $X_T$. [2 marks]

(B) The betweenness centrality of $X_S$ is always equal to that of $X_T$. [1 mark]

(C) The betweenness centrality of $X_S$ is always at most that of $X_T$. [2 marks]

(c) In directed graphs, the in-degree of a node $v$ is defined as the number of incoming edges $(u, v)$, whereas the node’s out-degree is defined as the number of outgoing edges $(v, u)$.

(i) What does high in-degree and out-degree mean in the context of the fishing collaboration? [2 marks]

(ii) Directed graphs are called “strongly connected” if there exists a path from every node to every other node. Is the graph in Figure 1 strongly connected? Justify your answer. [2 marks]

(iii) What is the relation between strong connectedness of a directed graph and its nodes’ in- and out-degrees? [3 marks]