2 Intermediacy

Consider a directed acyclic graph G = (V, E), where V denotes the set of nodes of G and E denotes the set of edges of G. The edges are directed. A certain node s \in V is referred to as the source, while another node t \in V is referred to as the target. We assume that each node v \in V is located on a path from the source s to the target t. We refer to such a path as a source-target path.

Definition 1

Given a source s and a target t, a path from s to t is called a \emph{source-target path}.

Informally, the more important the role of a node v \in V in connecting the source s to the target t, the higher the intermediacy of v. To define the concept of intermediacy in a formal way, we assume that each edge e \in E is active with a certain probability. We assume the probability of being active to be the same for all edges e \in E. This probability is denoted by p. The probability of an edge being inactive is denoted by q = 1 – p. Based on the idea of active and inactive edges, the following definitions can be introduced.

Definition 2

If all edges on a path are active, the path is called \emph{active}. Otherwise the path is called \emph{inactive}. If a node v \in V is located on an active source-target path, the node is called \emph{active}. Otherwise the node is called \emph{inactive}.

For two nodes u, v \in V, we use X\_uv to indicate whether there is an active path from node u to node v (X\_uv = 1) or not (X\_uv = 0). The probability that there is an active path from node u to node v is denoted by Pr(X\_uv = 1). We use X\_st(v) to indicate whether there is an active source-target path that goes through node v (X\_st(v) = 1) or not (X\_st(v) = 0). The probability that there is an active source-target path that goes through node v is denoted by Pr(X\_st(v) = 1). Note that this probability equals the probability that node v is active.

The concept of intermediacy can now be defined as follows.

Definition 3

The \emph{intermediacy} I(v) of a node v \in V equals the probability that v is active, that is, I(v) = Pr(X\_st(v) = 1).

2.1 Properties

To get a better understanding of the notion of intermediacy, we study the properties of intermediacy in two limit cases. We first consider the limit case in which the probability p of an edge being active tends to 0. We then consider the limit case in which the probability p tends to 1. It turns out that a ranking of the nodes in a graph based on intermediacy has a natural interpretation in each of these limit cases.

Let l\_v denote the length of the shortest source-target path going through node v \in V. Theorem 1 presented below states that in the limit as the probability p of an edge being active tends to 0, the ranking of nodes based on intermediacy coincides with the ranking based on l\_v, with nodes located on shorter source-target paths being more intermediate than nodes located on longer source-target paths.

The intuition for this result is as follows. When the probability of an edge being active is very close to 0, almost all edges are inactive. Consequently, almost all source-target paths are inactive as well. However, from a relative point of view, longer source-target paths are much more likely to be inactive than shorter source-target paths. This mean that, again from a relative point of view, nodes located on shorter source-target paths have a much higher probability to be active than nodes located on longer source-target paths (even though for all nodes this probability is close to 0). Nodes located on shorter source-target paths therefore have a higher intermediacy than nodes located on longer source-target paths.

Theorem 1

In the limit as p tends to 0, l\_u < l\_v implies I(u) > I(v).

Proof

Let m = |E| denote the number of edges in the graph G. Suppose that the m edges are split into two sets, one set of M edges and another set of m – M edges. The probability that the edges in the former set are all active while the edges in the latter set are all inactive equals

P\_M = p^M \* (1 – p)^(m – M).

Consider a node v \in V. The shortest source-target path that goes through v has a length of l\_v. This means that at least l\_v edges need to be active in order to obtain an active source-target path that goes through v. Hence, the probability that there is an active source-target path that goes through v can be written as

I(v) = sum\_{i = l\_v}^m n\_vi \* P\_i,

where n\_vi > 0 for all i = l\_v, ..., m. Note that this probability equals the intermediacy of v.

Now consider two nodes u, v \in V with l\_u < l\_v. In the limit as p tends to 0, I(u) and I(v) both tend to 0. However, they do so at different rates. More specifically, in the limit as p tends to 0, we have

I(v) / I(u)

= [sum\_{i = l\_v}^m n\_vi \* P\_i] / [sum\_{i = l\_u}^m n\_ui \* P\_i]

= [sum\_{i = l\_v}^m n\_vi \* P\_i / P\_{l\_u}] / [sum\_{i = l\_u}^m n\_ui \* P\_i / P\_{l\_u}]

= [sum\_{i = l\_v}^m n\_vi \* p^{i – l\_u} \* (1 – p)^{l\_u – i}] / [sum\_{i = l\_u}^m n\_ui \* p^{i – l\_u} \* (1 – p)^{l\_u – i}]

= 0 / n\_ui

= 0

Hence, in the limit as p tends to 0, I(u) > I(v).

Let k\_v denote the number of edge independent source-target paths going through node v \in V. Theorem 2 presented below states that in the limit as the probability p of an edge being active tends to 1, the ranking of nodes based on intermediacy coincides with the ranking based on k\_v. The larger the number of edge independent source-target paths going through a node, the higher the intermediacy of the node.

The intuition for this result is as follows. When the probability of an edge being active is very close to 1, almost all edges are active. Consequently, almost all source-target paths are active as well, and so are almost all nodes. A node is inactive only if all source-target paths going through the node are inactive. If there are k edge independent source-target paths that go through a node, this means that the node can be inactive only if there are at least k inactive edges. Consider two nodes u and v. Suppose that the number of edge independent source-target paths going through node u is larger than the number of edge independent source-target paths going through node v. In order to be inactive, node u then requires a larger number of inactive edges than node v. Since the probability of an edge being active is very close to 1, this means that, from a relative point of view, node u has a much lower probability of being inactive than node v (even though for both nodes this probability is close to 0). Hence, node u has a higher intermediacy than node v. More generally, nodes located on a larger number of edge independent source-target paths have a higher intermediacy than nodes located on a smaller number of edge independent source-target paths.

Theorem 2

In the limit as p tends to 1, k\_u > k\_v implies I(u) > I(v).

Proof

Let m = |E| denote the number of edges in the graph G. Suppose that the m edges are split into two sets, one set of M edges and another set of m – M edges. The probability that the edges in the former set are all inactive while the edges in the latter set are all active equals

Q\_M = q^M \* (1 – q)^(m – M).

Consider a node v \in V. There are k\_v edge independent source-target paths that go through v. This means that at least k\_v edges need to be inactive in order for there to be no active source-target path that goes through v. Hence, the probability that there is no active source-target path that goes through v can be written as

R\_v = sum\_{i = k\_v}^m n\_vi \* Q\_i,

where n\_vi > 0 for all i = k\_v, ..., m. Note that the intermediacy of v equals 1 minus the above probability, that is, I(v) = 1 – R\_v.

Now consider two nodes u, v \in V with k\_u > k\_v. In the limit as p tends to 1, R\_u and R\_v both tend to 0. However, they do so at different rates. More specifically, in the limit as p tends to 1, we have

R\_u / R\_v

= [sum\_{i = k\_u}^m n\_ui \* Q\_i] / [sum\_{i = k\_v}^m n\_vi \* Q\_i]

= [sum\_{i = k\_u}^m n\_ui \* Q\_i / Q\_{k\_v}] / [sum\_{i = k\_v}^m n\_vi \* Q\_i / Q\_{k\_v}]

= [sum\_{i = k\_u}^m n\_ui \* q^{i – q\_v} \* (1 – q)^{q\_v – i}] / [sum\_{i = k\_v}^m n\_vi \* q^{i – q\_v} \* (1 – q)^{q\_v – i}]

= 0 / n\_vi

= 0

Hence, in the limit as p tends to 1, R\_u < R\_v, which implies that I(u) > I(v).

Theorems 1 and 2 are concerned with the properties of intermediacy in the limit cases in which the probability p of an edge being active tends to either 0 or 1. Figure 1A provides some insight into the behavior of intermediacy for values of the probability p that are in between these two extremes. The figure shows two graphs. In the left graph, there is a direct path from node u to node v. There are no indirect paths. In this graph, the probability that there is an active path from u to node v equals p. In the right graph, there is no direct path from node u to node v. However, there are k indirect paths of length 2. Each of these paths has a probability of p^2 of being active. Consequently, the probability that there is at least one active path from node u to node v equals 1 – (1 – p^2)^k. The table shown in Figure 1A presents for different values of k the value of p for which the probability that there is an active path from node u to node v is the same in the left and the right graph. For instance, suppose that k = 5 and p = 0.22. Clearly, in the left graph, the probability that there is an active path from node u to node v then equals 0.22. However, in the right graph, this probability also equals 0.22, since 1 – (1 – p^2)^k = 0.22. In other words, when the probability p of an edge being active is set to 0.22, a direct path between two nodes is considered equally strong as 5 indirect paths of length 2. Likewise, the table for instance indicates that for p = 0.62 a direct path between two nodes is considered equally strong as 2 indirect paths of length 2. Based on Figure 1A, one can choose a value for the probability p that one considers appropriate for a particular analysis.