Betweenness Centrality

Definition:

$$b(v, G) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

$$b_n(v, G) = \sum_{s \neq v \neq t} \frac{1}{n(n-1)} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where $\sigma_{st}(v)$ is the number of shortest paths from s to t with v in it and $\sigma_{st}$is the number of shortest paths from s to t. n is the number of nodes in the graph. For large graphs an algorithm to approximate $b_n(v, G)$ is

- Set count = 0
- For each $i \in \{1, \ldots, K\}$:
  - Pick $s \leftarrow$ at random from V
  - Pick $t \neq s \leftarrow$ at random from V
  - Run: all shortest paths from s to t
  - $\Pi_s \leftarrow$ shortest path from s to t at random
  - if $v \in \Pi_{st}$:
    - Set count = count + 1
- return count/K

Proposition:

$$\lim_{K \to \infty} \frac{1}{K} \sum_{i=1}^{K} \text{count} \approx b(v, G)$$

Proof: Let

$$X_i = \begin{cases} 1 & \text{if } v \in \Pi_{i,st} \\ 0 & \text{else} \end{cases}$$

$$\frac{1}{K} \sum_{i=1}^{K} X_i = \frac{\text{count}}{K}$$

$$E(\text{count}) = E(\sum_{i=1}^{K} X_i) = \sum_{i=1}^{K} E(X_i) = \sum_{i=1}^{K} Pr(X_i = 1)$$

$$= \sum_{i=1}^{K} \sum_{s,t \in V; s \neq t} Pr(X_i = 1 | s, t) Pr(s, t) = \sum_{i=1}^{K} \sum_{s,t \in V; s \neq t} Pr(X_i = 1 | s, t) \frac{1}{n(n-1)}$$

$$= \sum_{i=1}^{K} \sum_{s,t \in V; s \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \frac{1}{n(n-1)} = K * \sum_{s,t \in V; s \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \frac{1}{n(n-1)} = K * b_n(v, G)$$

Thus

$$\hat{b}(v, G) = \frac{1}{n(n-1)} \sum_{i=1}^{K} \frac{\text{count}}{K} = b(v, G)$$