Inclusion probabilities

- For incident subgraph sampling, edge inclusion probabilities are

$$_{\{i;j\}} = \frac{n}{N_e}$$

- Vertex in V if any one or more of its incident edges are sampled

$$\begin{aligned} {}_i &= P(\text{vertex } i \text{ is sampled}) \\ &= 1 \quad P(\text{no edge incident to } i \text{ is sampled}) \\ &= \begin{cases} 1 \quad \dfrac{\binom{N_e \quad d_i}{n}}{\binom{N_e}{n}}; & \text{if } n \quad N_e \quad d_i \\ 1; & \text{if } n > N_e \quad d_i \end{cases} \end{aligned}$$

- Vertices included with unequal probs. that depend on their degrees
  - Probability proportional to size (degree) sampling of vertices
  - Requires knowledge of N_e and degree sequence $\{d_i\}_{i \in V}$

Example: Estimating network size

- Consider estimating N_e as an edge total, i.e.,

$$N_e = \sum_{(i;j) \in E} 1 = \sum_{(i;j) \in V^{(2)}} A_{ij}$$

Example: Estimating network size (cont.)

- Protein network: $N_v = 5;151$, $N_e = 31;201$
  - BS of vertices with $p = 0:1$ and $p = 0:3$
  - Process repeated for 10,000 trials; histogram of $\hat{N}_e$

- Average of $\hat{N}_e$ was 31,116 and 31,203 Unbiasedness supported
  - Mean and variability of $\hat{N}_e$ shrinks with p (larger sample)