ADAPTIVE WEIGHTING ESTIMATION ON A BIPARTITE GRAPH

MARTINA PATONE
LI-CHUN ZHANG

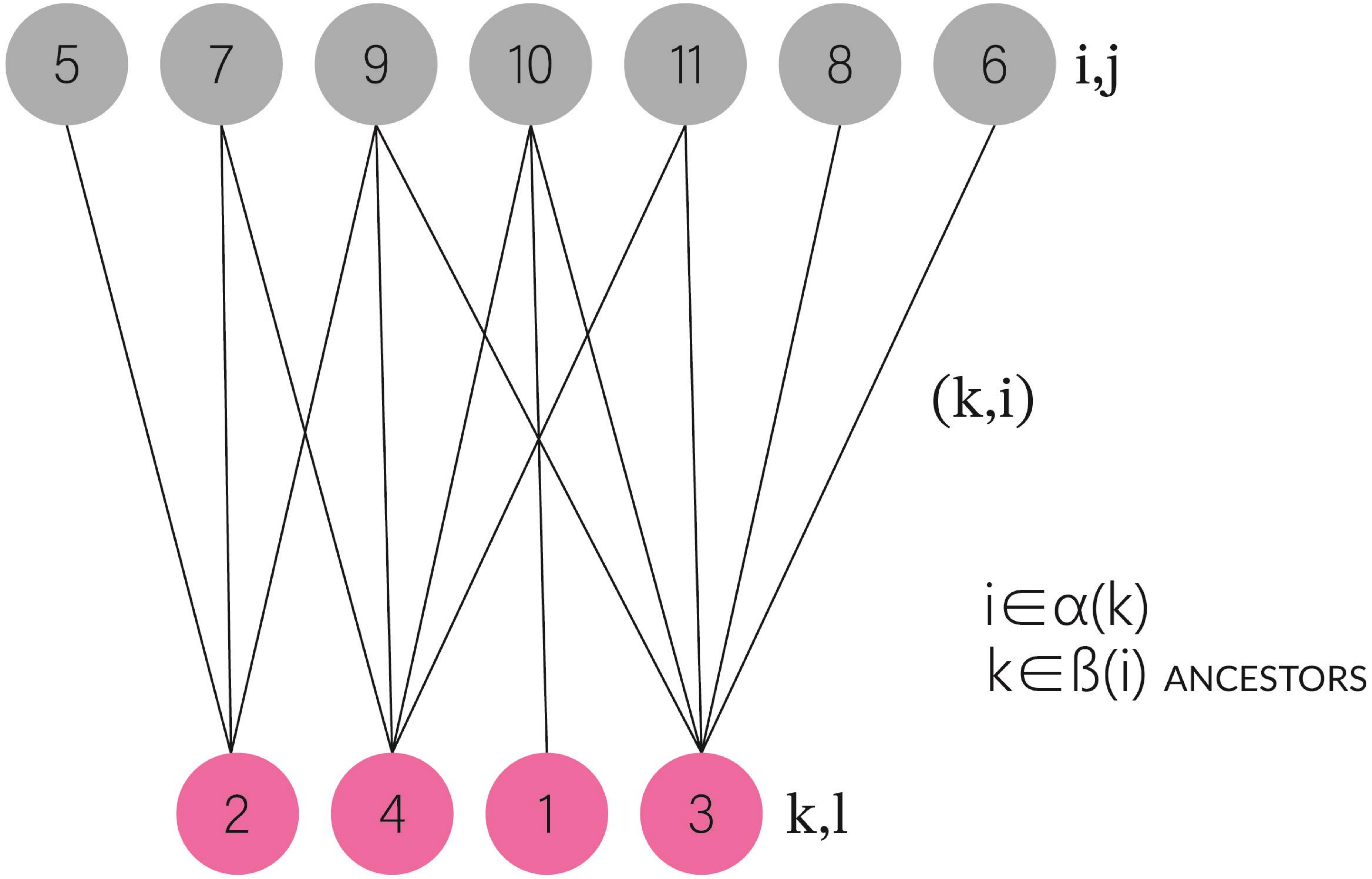
Web Science
Centre for Doctoral Training

dstl

UNIVERSITY OF
Southampton

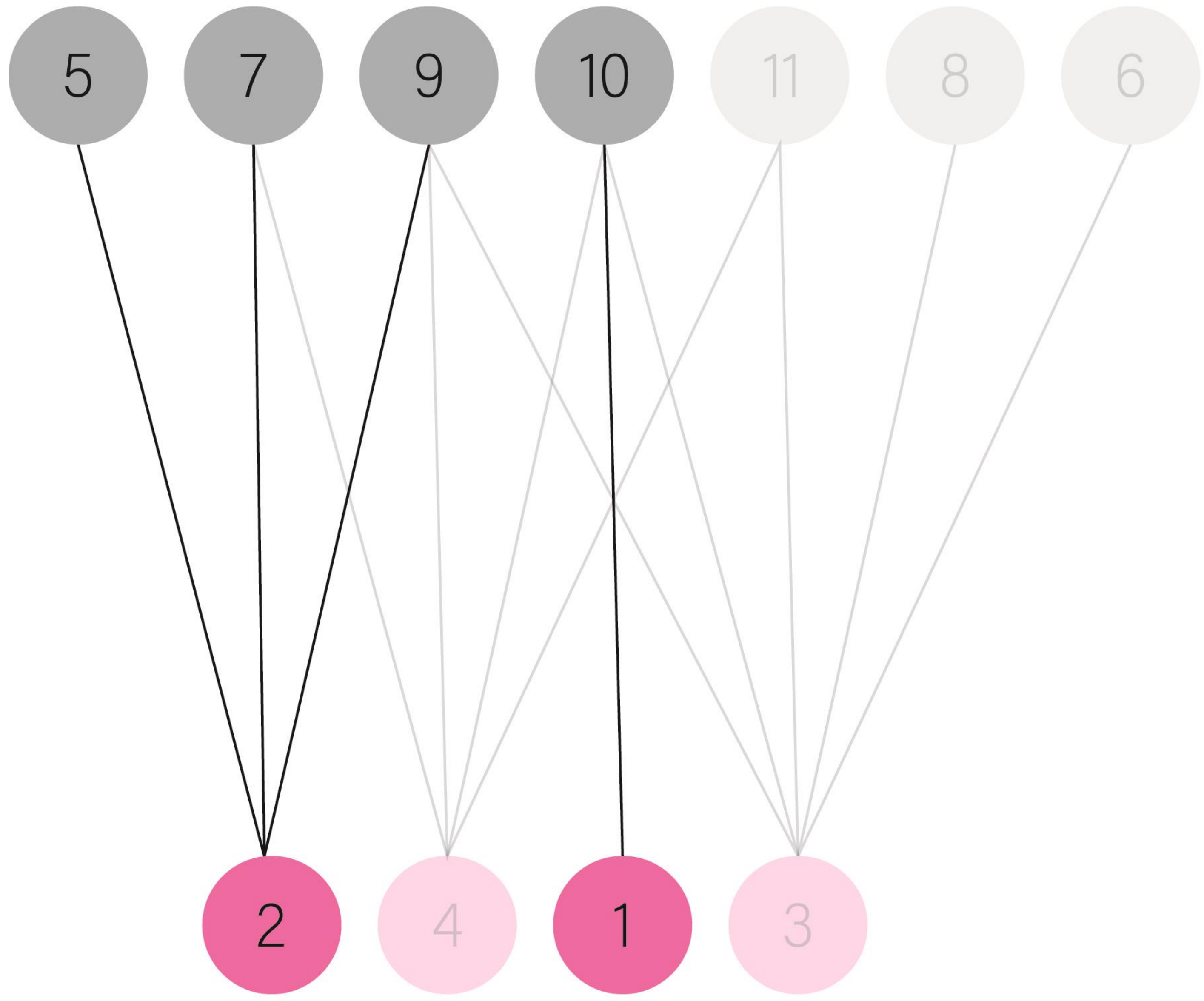

THE 'MULTIPLICITY' PROBLEM

Population graph $G = (F, U; A)$,
F: sampling frame (pink); U: population of motifs (grey).

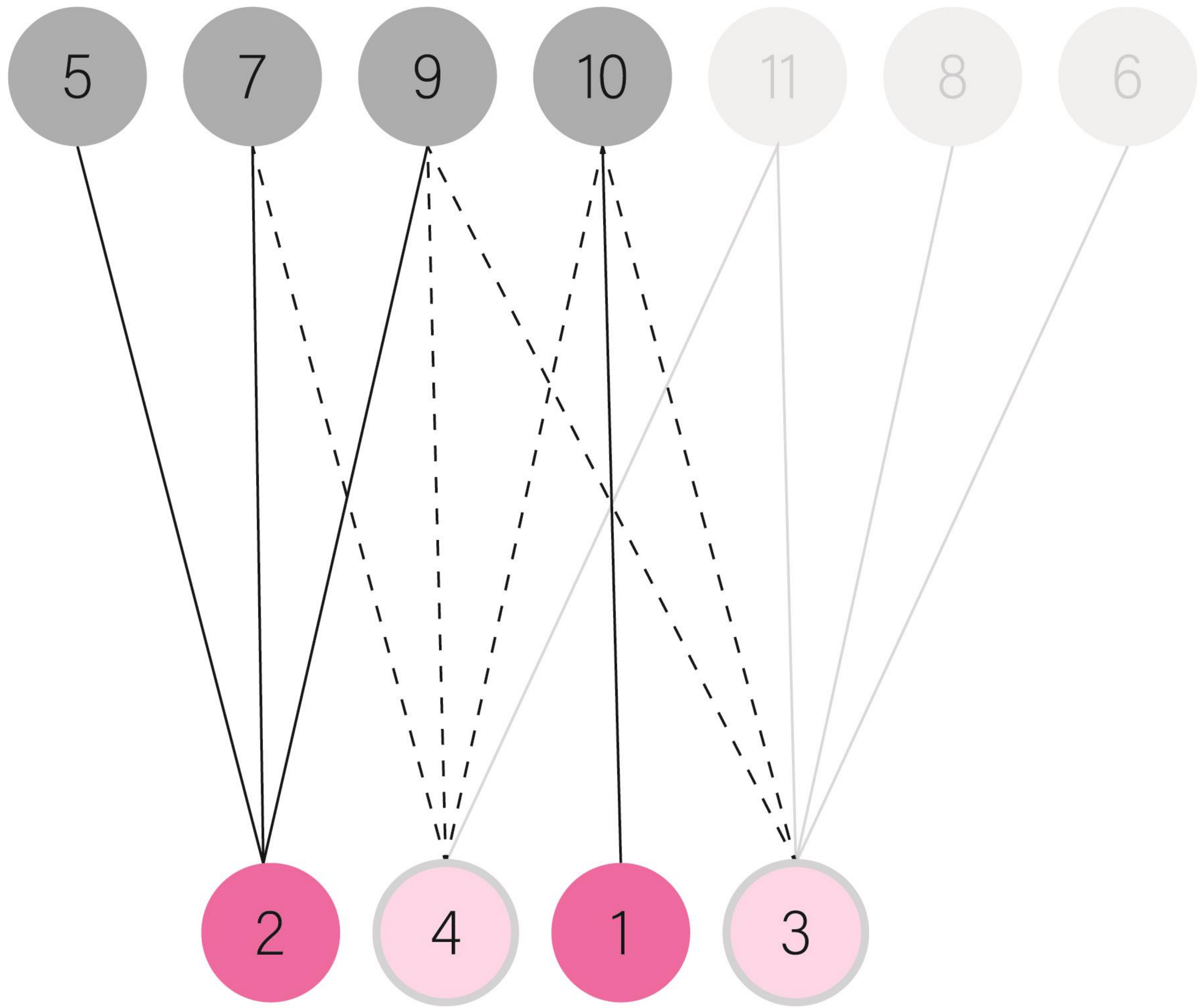


Bipartite Incident Graph.

Sample BIG, $G_s = (s \cup \alpha(s); A_s)$,

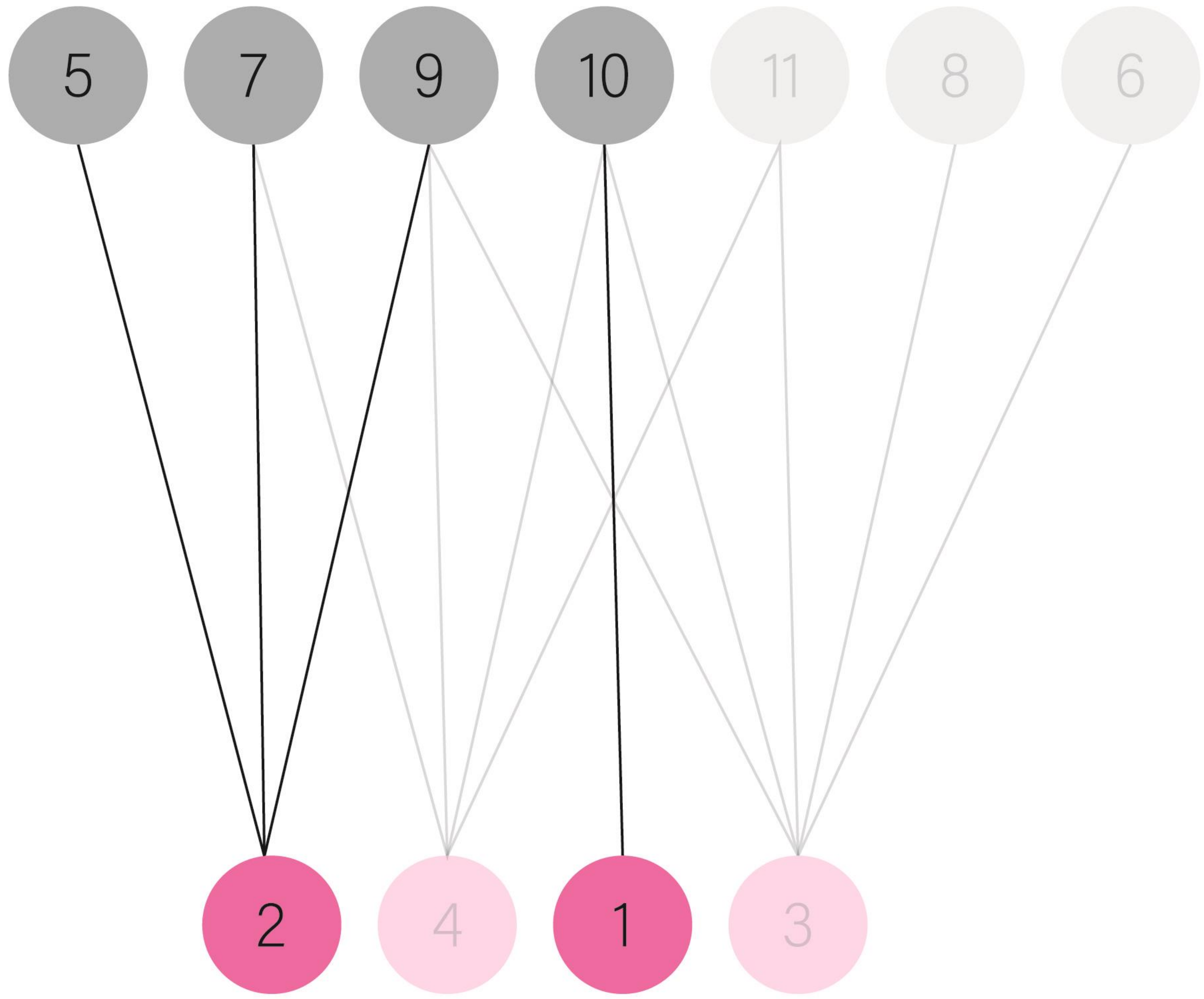


Incident reciprocal observation.

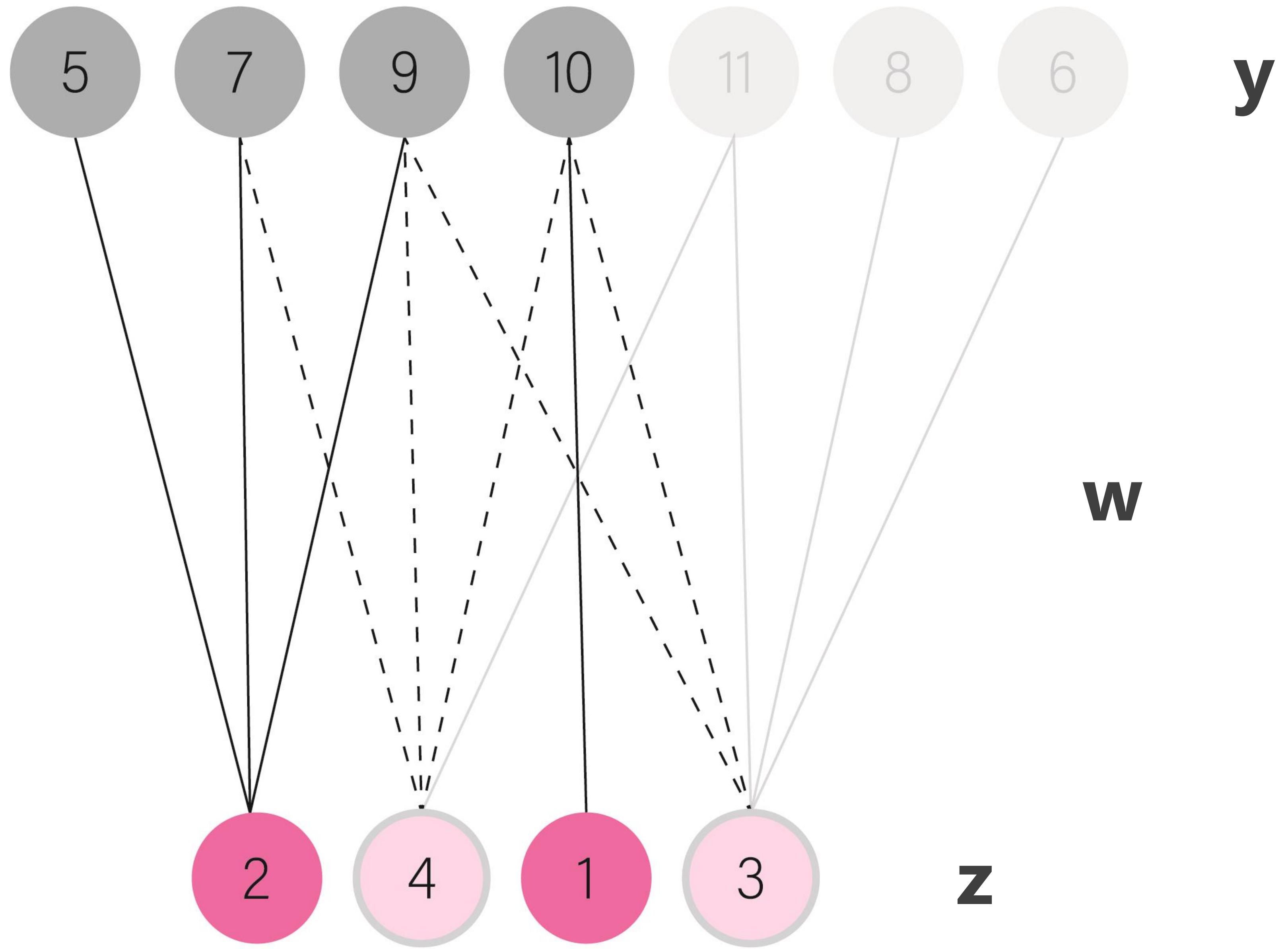


Incident ancestral observation.

Sample BIG, $G_s = (s \cup \alpha(s); A_s)$,



Incident reciprocal observation.



Incident ancestral observation.

Let θ be the parameter of interest. We have:

$$\theta = \sum_{i \in U} y_i = \sum_{k \in F} z_k = \sum_{(ki) \in A} w_{ki} y_i \ ,$$

where z_k is a constructed measure for each unit in F, which is given by

$$z_k = \sum_{i \in \alpha(k)} w_{ki} y_i \qquad \text{and} \qquad \sum_{k \in \beta(i)} w_{ki} = 1$$

Yhat: (Horvitz and D. J. Thompson, 1952)

$$\hat{\theta}_y = \sum_{i \in \alpha(s)} \frac{y_i}{\pi_{(i)}} = \sum_{i \in U} \frac{\delta_{(i)}}{\pi_{(i)}} y_i \ ;$$

Zhat: (Birnbaum and Sirken, 1965)

$$\hat{\theta}_z = \sum_{k \in s} \sum_{i \in \alpha(s)} \frac{w_{ki} y_i}{\pi_k} = \sum_{k \in F} \frac{\delta_k}{\pi_k} z_k \ ;$$

Phat: (Birnbaum and Sirken, 1965)

$$\hat{\theta}_p = \sum_{(ki) \in A_s} \frac{I_{ki} w_{ki}}{p_{(ki)}} \cdot \frac{y_i}{\pi_k} = \sum_{k \in s} \sum_{i \in \alpha(s)} \frac{W_{ki} y_i}{\pi_k} = \sum_{k \in s} \frac{Z_k}{\pi_k} \ .$$

$W_{ki} = \frac{w_{ki} I_{ki}}{p_{(ki)}}$, with $I_{ki} = 1$ if $k = \min(\beta(i) \cap s)$.

Zhat: $w_{ki} = 1/d_i$;

Phat: $W_{ki} = \frac{w_{ki} I_{ki}}{p_{(ki)}}$, with $I_{ki} = 1$ if $k = \min(\beta(i) \cap s)$.

Let $i = 10$, $d_{10} = 3$. Assume SRS of size 2.

$$\left(p_{(1,10)}, p_{(3,10)}, p_{(4,10)}\right) = \left(1, \frac{2}{3}, \frac{1}{3}\right) \ .$$

Equal-share weights w and priority weights W for the edges incident to $i = 10$.

s	$w_{1,10}$	$w_{3,10}$	$w_{4,10}$
$\{1,3\}$	$1/3$	$1/3$	-
$\{3,4\}$	-	$1/3$	$1/3$
$\{2,4\}$	-	-	$1/3$

s	$W_{1,10}$	$W_{3,10}$	$W_{4,10}$
$\{1,3\}$	$1/3$	0	-
$\{3,4\}$	-	$1/2$	0
$\{2,4\}$	-	-	1

Let the *adaptive weight* be given by

$$W_{ki} = h(w, s, t) \qquad \text{for } (ki) \in A_s \ ,$$

where w is the set of initial fixed weights and t denotes generically the auxiliary information that is extraneous to G_s.

The *adaptive weighting estimator (AWE)* based on W is given by

$$\hat{\theta}_A = \sum_{k \in s} \frac{Z_k}{\pi_k} = \sum_{(ki) \in A_s} \frac{W_{ki} y_i}{\pi_{(ki)}} \ ,$$

where $\pi_{(ki)} = \pi_k$.

Proposition 1. The AWE is unbiased for θ provided, for each $i \in U$,

$$\sum_{k \in \alpha(i)} E(W_{ki} | \delta_k = 1) = 1 \ .$$

Proposition 2. The variance of an unbiased AWE can be given by

$$V(\hat{\theta}_A) = V(\hat{\theta}_z) + \Delta$$

where $V(\hat{\theta}_z)$ is the variance of Zhat based on the initial weights w, and

$$\Delta = \sum_{k \in F} \sum_{l \in F} \frac{\pi_{kl}}{\pi_k \pi_l} \sum_{i \in \alpha(k)} \sum_{j \in \alpha(l)} \Big(E(W_{ki} W_{lj} | \delta_k \delta_l = 1) - w_{ki} w_{lj} \Big) y_i y_j \ .$$

$$W_{ki} = h(w, s, t)$$

By Prioritisation: let $t = I_{ki}$ and $p_{(ki)} = \mathrm{Pr}(I_{ki} = 1 | \delta_k = 1)$,

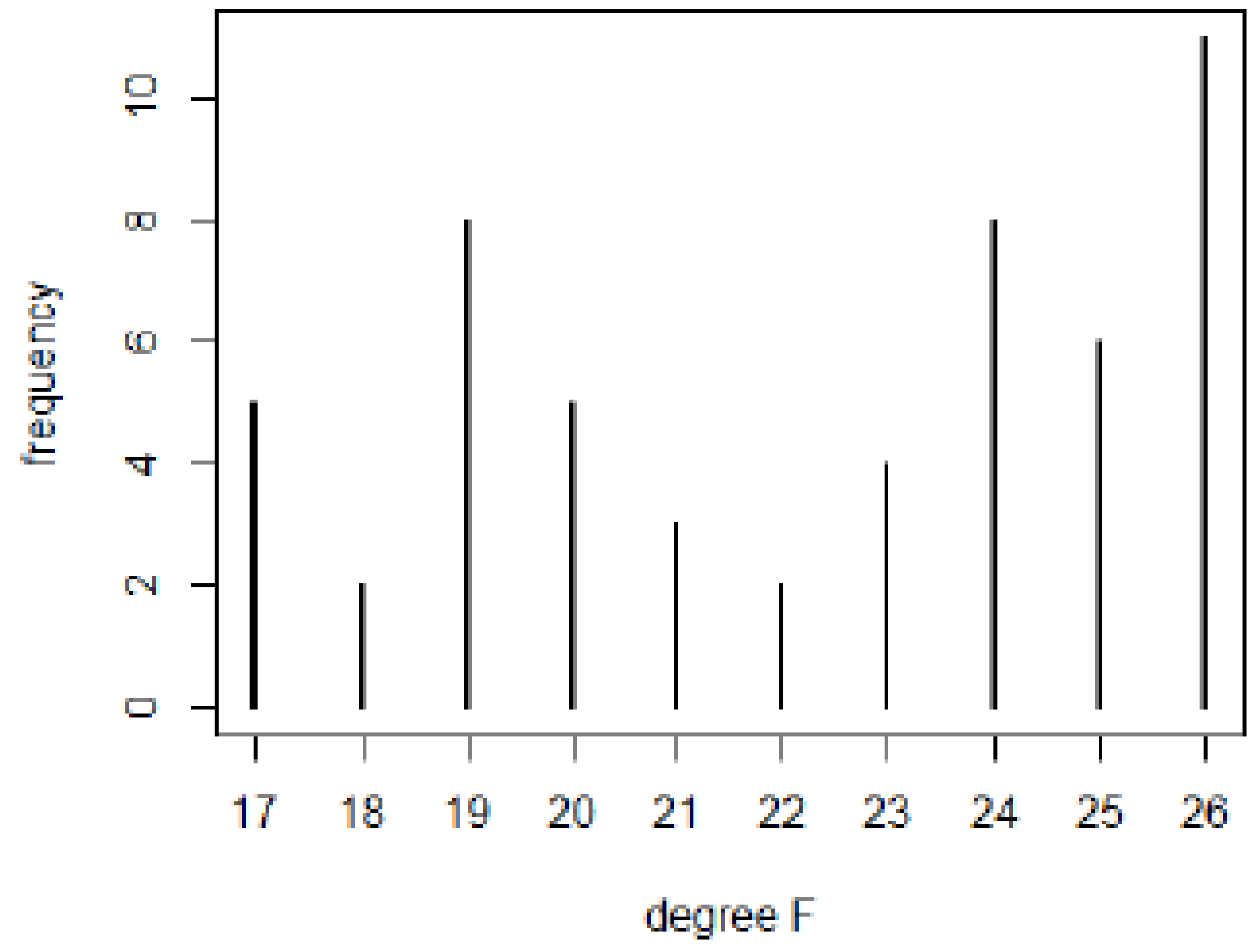
$$W_{ki} = \frac{w_{ki} I_{ki}}{p_{(ki)}} \ ;$$

By Resharing: let $t = g_{(ki)}$,

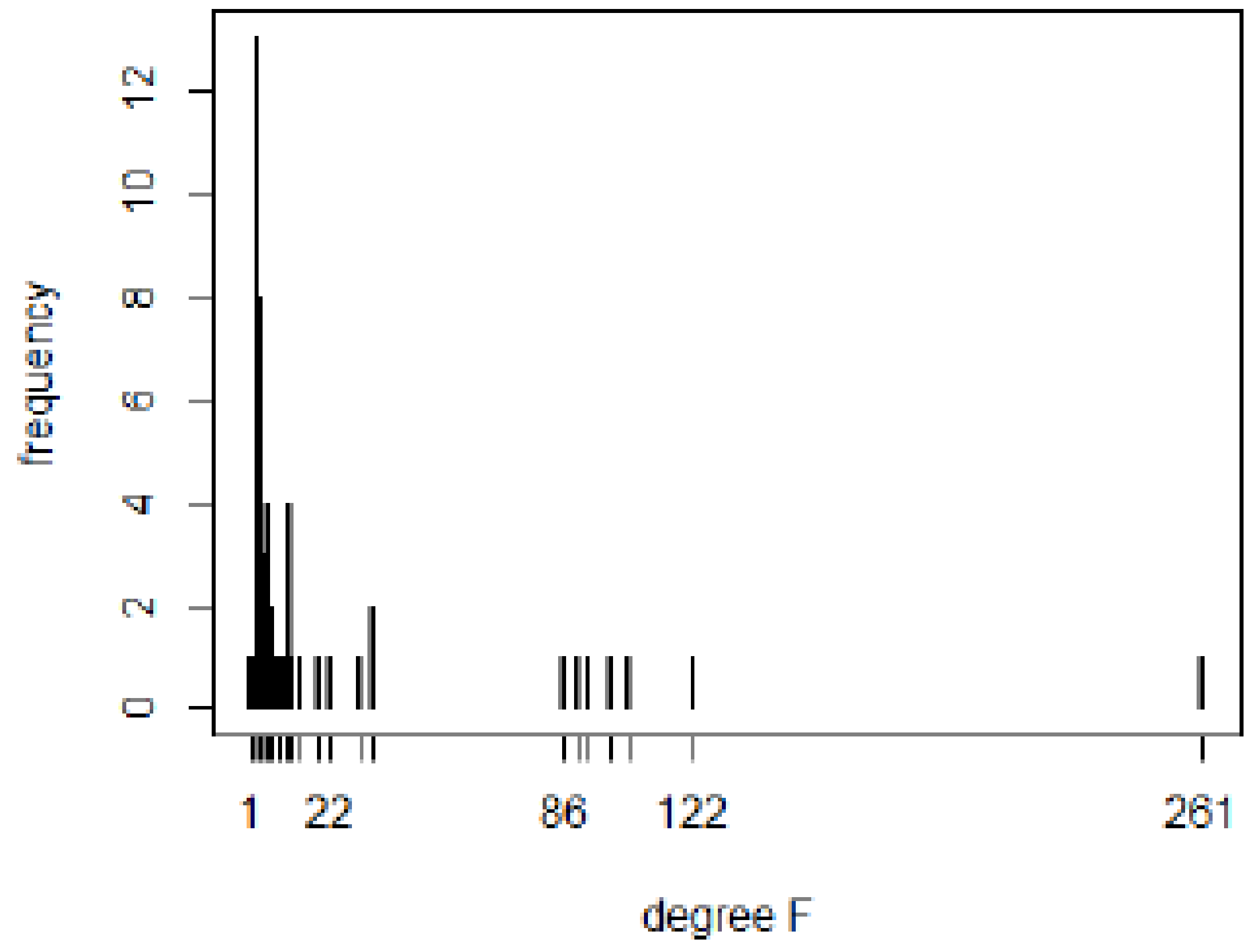
$$W_{ki} = w_{ki} g_{(ki)} \ .$$

Let $G_1 = (F \cup U, A_1)$ and $G_2 = (F \cup U, A_2)$.
$|F| = 54$ and $|U| = 310$.

Assume SRS of size m from F. Let $\theta = |U|$.



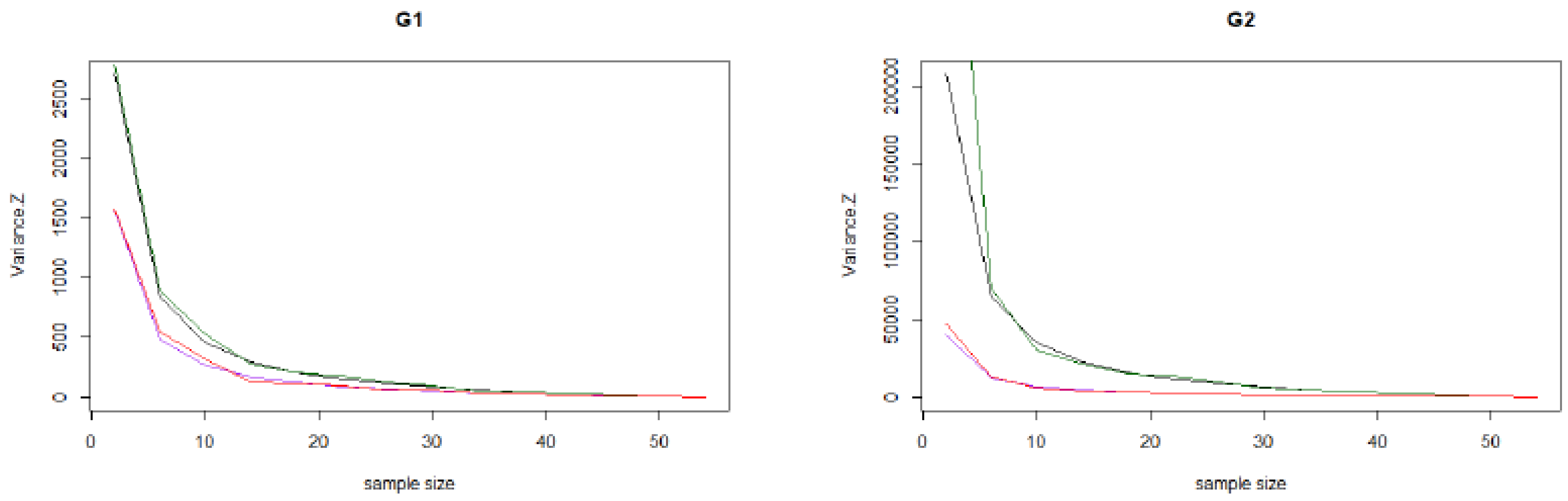
Degree distribution for the sampling units in G_1.



Degree distribution for the sampling units in G_2.

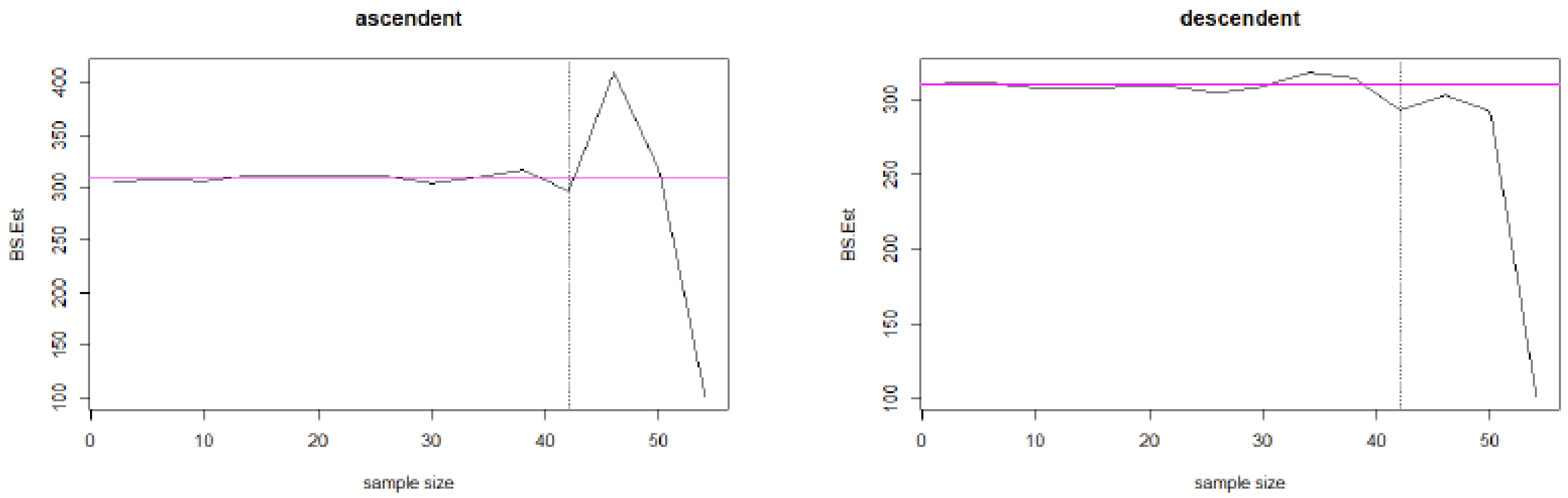
$$W_{ki} = h(w, s, t) = \text{id}(w) = w_{ki}$$

1. Equal-share weights (Birnbaum and Sirken, 1965): $w_{ki} = \frac{1}{d_i}$;
2. Unequal-share weights: $w_{ki} = \frac{1}{d_k} / \sum_{l \in \beta(i)} \frac{1}{d_l}$.



True and estimated variance of the AWE over 100 simulations for two different choices of fixed weights: (green-black) - equal share; (red-purple) - unequal share.

Phat: $W_{ki} = \frac{w_{ki} I_{ki}}{p_{(ki)}}$, with $I_{ki} = 1$ if $k = \min(\beta(i) \cap s)$.



Average of the Priority AWE (BS, 1965) over 100 simulations for the graph G_1 with different ordering of the frame.

The *bound for unbiasedness* is the maximum m which guarantees that, for any motif $i \in U$:

$$\Pr[k, l \in s | m] < 1, \text{ where } k, l \in \beta(i) \text{ with } k \neq l \ .$$

BIG + AWE :

GENERAL AND FLEXIBLE SAMPLING STRATEGY

NEW INSIGHTS ON THE EXISTING ESTIMATORS

MORE EXAMPLES OF AWE

GAIN IN EFFICIENCY WITH DIFFERENT CHOICES OF WEIGHTS

THANKS