THE SUPPLEMENTARY MATERIAL FOR

"A NEW PREFERENTIAL MODEL WITH HOMOPHILY FOR RECOMMENDER SYSTEMS"

Hanyang Tian, Bo Zhang, Ruixue Jiang, and Xiao Han

University of Science and Technology of China

Abstract: This supplementary material contains the applicability of KPA models, detailed introduction to the estimators' robustness, and the group label recovery in Section 6 of our main paper, the basic information of datasets in Section 8, and the proofs of lemmas and theorems in Sections 3–5.

S1. Applicability of KPA models

We generate three network datasets from different models:

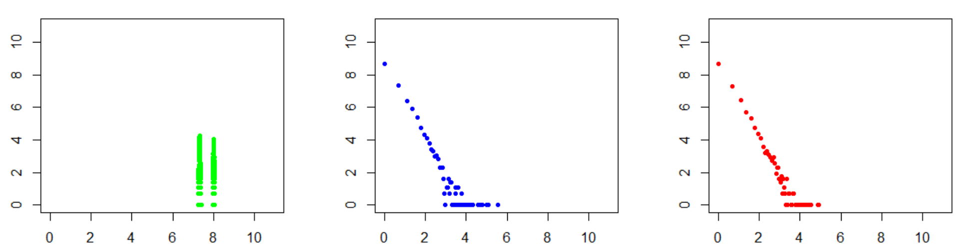
a. by a stochastic block model with $N = 10000$ nodes, $K = 2$ groups, the corresponding membership vector $g = (\underbrace{1, 1, \cdots}_{5000}, \underbrace{2, 2, \cdots}_{5000})$, and the group connection probability matrix $B = \begin{pmatrix} 0.5 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}$.

b. by an original preferential attachment model on the time range $[0, 100000]$, the vertex-step probability $q = 0.5$. The initial graph $G(0)$ has one node with a loop.

c. by a KPA model on the time range $[0, 100000]$, vertex-step probability $q = 0.5$, $K = 2$ groups, group probability $p = (0.5, 0.5)$, and homophily parameter $\theta = 0.5$. The initial graph $G(0)$ has 5 isolated nodes from group 1 and 5 isolated nodes from group 2. They each have a loop.

To judge whether the KPA model is suitable for a specific network data set, we need to test the existence of rich-get-richer and homophily of the data set.

- The preferential attachment model's embodiment of the rich-get-richer mechanism is the power-law distribution of nodes' degrees. Thus, we can identify the rich-get-richer by the nodes' power-law distribution. Let $F(d)$ record the number of nodes with degree d. The degree distribution is plotted as follows:



(a) SBM (b) PAM (c) KPAM

Figure S1: The x-axis is the $\log(d)$, the y-axis is the $\log(F(d))$.

Figure S1 implies that the network without the rich-get-richer mechanism does not exhibit the power-law property.

Thus, when faced with an unknown network data set, we can first make the degree distribution plot of the network to judge whether it has a rich-get-richer mechanism. If not, which does not fit our KPA model. Otherwise, we can further test the mechanism of homophily.

- Since we determine that the network data has a power-law degree distribution, we can assume that a family of preferential attachment models generates the network. Further, we can use the method in Section 4 of our main paper to estimate the parameter θ and test the hypothesis on homophily.

S2. Robustness

S2.1 Erroneous group labels

The absence of group labels for nodes can be considered a special case of erroneous group labels since a random label can be assigned to each node without a label.

The nodes' group labels are essential for the MLE and snapshot estimation. However, some nodes may have labels that do not match their actual grouping structure. Therefore, we need to examine the impact of erroneous labeling on the convergence of our estimators.

First, we consider the case with $\theta^* = 1$.

Assumption S.1. The added node w from group k_1 in the vertex-step has a probability p_{k_1,k_2} of being labeled as $g_w = k_2$, where $k_2 \neq k_1 \in [K]$ and for any $k_1 \in [K]$, $\sum_{k_2,\, k_2 \neq k_1} p_{k_1,k_2} \in [0, 1)$.

Theorem S.1. *Suppose an evolving network is generated by a KPA model with $\theta^* = 1$, and the added nodes satisfy Assumption S.1. For this network data, mislabeling does not affect our MLE of parameter θ^*. Our estimator still converges to 1, and the CLT still holds:*

$$\hat{\theta} \xrightarrow{a.s.} 1;$$
$$T^{1/2}(\hat{\theta} - 1) \xrightarrow{d} N(0, \frac{1}{K-1}),$$

when T tends to infinity.

Theorem S.1 ensures that we can still infer the network has a homophily structure when $\hat{\theta} < \theta_\alpha$ by criterion (4.5) in the main paper, even if the network has a large probability of mislabeling.

Now we consider the second case: $\theta^* < 1$. We test our estimators' robustness in Simulations S3.1–S3.2 with some nodes being mislabeled.

The results in Tables S1–S2 imply that the impact on the estimation is slight when mislabeled nodes occur with a small probability. However, Tables S3–S4 show that if the mislabeled nodes have high degrees, even in small and finite numbers, they can severely impact the parameter estimation.

If we can recover the group labels of the nodes with high degrees, the effectiveness of our estimation methods will be significantly improved. Group label recovery (community recovery) is a hot topic in network research. Hajek and Sankagiri (2019) proposed two algorithms to recover the community labels of finite nodes in preferential attachment models.

For node u in $G(T)$, let:

$$\mathcal{C}_u = \{w \in \mathcal{V}(T) : \exists t \in [1, T], e(t) = (w, u) \text{ and } v(t) = 1\}.$$

We call $\mathcal{C}_u$ the children set of node u. Inspired by Hajek and Sankagiri (2019), we propose a method to recover the group label of a node based on its children.

Let

$$\hat{g}_u = \arg\max_{k \in [K]} \frac{\sum_{i \in \mathcal{C}_u} \mathbb{1}\{g_i = k\}}{\hat{p}_k}, \; \hat{p}_k = \frac{|\mathcal{V}_k(T)|}{|\mathcal{V}(T)|}. \tag{S2.1}$$

where $\mathcal{V}_k(T)$ is the set of nodes from group k at time T.

Theorem S.2. *Under the conditions of Theorem 1 in the main paper and with $\theta^* < 1$, suppose that only a finite number of nodes are mislabeled and* *we know the group labels of nodes in* $\mathcal{C}_u$. *Then the estimated label* $\hat{g}_u$ *by Equation (S2.1) satisfies*

$$\hat{g}_u \xrightarrow{i.p.} g_u, \text{ with } |\mathcal{C}_u| \to \infty.$$

Table S7 shows the effectiveness of this method.

S2.2 Missing of observation

We will discuss the missing cases of edge connection in vertex-steps and edge-steps in detail.

- Edges that connect nodes are missing in vertex-steps. In a vertex-step, a new node arrives with an edge connecting to an existing network node. If the edge in the vertex-step is missing, we cannot observe any edges that connect to this new node until it connects again to the existing nodes without missing. Missing edges in vertex-steps can result in a significant discrepancy between the observed and the actual network structure.

- Edges that connect nodes are missing in edge-steps. In an edge-steps, a new edge appears between two existing nodes in the network without any new nodes arriving. The missing observation for this edge results in an underestimation of the network size by one edge.

We design Simulation S3.3 to verify our estimators' convergence when some edges are missed. Based on Tables S5–S6, it can be inferred that MLE is more robust than snapshot estimation in the presence of missing observations at edges.

S2.3 Instability

Under Assumption 3 in the main paper, the probability of a vertex-step occurring is constant. However, in real-world network data, the frequency of nodes entering the network may not remain constant. Therefore, this subsection discusses whether our estimation method remains effective when the parameter q undergoes a finite number of changes within the time range $[0, T]$.

Assumption S.2. Parameter $q \in (0, 1)$ changes at time $\{\tau_l\}_{l=1}^{L}$, where for each l, $0 < \tau_l < T$, $\tau_l / T = c_l \in (0, 1)$.

Set

$$q = \begin{cases} q_1, & \text{if } t \leq \tau_1; \\ q_2, & \text{if } \tau_1 < t \leq \tau_2; \\ \vdots & \\ q_{L+1}, & \text{if } \tau_L < t \leq T. \end{cases}$$

Further, let $\tau_0 = 0$, $\tau_{L+1} = T$; $c_0 = 0$, $c_{L+1} = 1$.

Theorem S.3. *Under Assumptions 1–3 in the main paper and Assumption S.2, let $d_i(t)$ be the degree of node i in graph $G(t)$. Let $D_k(t) = \sum_{i\in\mathcal{V}(t)} d_i(t)\mathbb{1}\{g_i = k\}$ be the total degrees from group k in $G(t)$, for $k \in [K]$. When T tends to infinity,*

$$\frac{D_k(\tau_l)}{2\tau_l} \xrightarrow{a.s.} p_k, \quad l = 1, \cdots, L+1.$$

Theorem S.4. *Assume the evolving network $\{G(t)\}_{t=0}^{T}$ is generated by a KPA model under Assumptions 1–3 in the main paper and Assumption S.2. For the true parameter θ^*, $\hat{\theta}$ is the MLE and $\tilde{\theta}$ is the snapshot estimation. When T tends to infinity,*

$$\hat{\theta}, \tilde{\theta} \xrightarrow{a.s.} \theta^*.$$

Theorems S.3–S.4 guarantee the convergence of our estimation method even when the parameter q undergoes a finite number of changes. Moreover, we design Simulation S3.5 to verify the convergence of estimators $\hat{\theta}$ and $\tilde{\theta}$ when q changes.

S3. Simulation

S3.1 Simulation 1 for Erroneous group labels

This subsection tests the estimators' robustness when mislabeled nodes are added with a small probability. Set $B = 500$, $T \in \{100, 500, 2500, 5000\}$, $n_0 = 10$, $q = 0.5$, $p = (\underbrace{0.1, 0.1, \cdots}_{10})$.

Further, in a vertex-step, the new node from group k arrives with probability p_k, and its observed group label has a probability p_w of being a random sample from $[K]$, $p_w \in \{\log(T)/T, \log^2(T)/T\}$.

Set $\{g_i\}_{i=1}^n$ be the nodes' true group labels and $\{D_k(t)\}_{k=1}^K$ be the true groups' degrees at time t. And we use $\{\breve{g}_i\}_{i=1}^n$, $\{\breve{D}_k(t)\}_{k=1}^K$ to record the observed group labels and groups' degrees.

$$\breve{g}_i = \begin{cases} g_i, & \text{with probability } 1 - p_w; \\ \text{a sample from } [K], & \text{with probability } p_w. \end{cases}$$

$\{\{\breve{D}_k(t)\}_{k=1}^K, v(t), \breve{g}_{e_1(t)}, \breve{g}_{e_2(t)}\}_{t=1}^T$ are used to construct the maximum likelihood estimation (Equation (4.1) in the main paper). Record the estimator of θ as $\hat{\theta}_{\text{mle}}(b)$ in the bth trial, $b \in [B]$. $\{\breve{\mathcal{E}}_{k,0}(T)\}_{k=1}^K$ record the number of edges from group k to another and $\{\breve{\mathcal{E}}_{k,1}(T)\}_{k=1}^K$ record the number of edges that both nodes from group k under the observed group labels. $\{\breve{D}_k(T), \breve{\mathcal{E}}_{k,1}(T), \breve{\mathcal{E}}_{k,0}(T)\}_{k=1}^K$ are used to construct the snapshot estimation

(Equation (4.9) in the main paper). Record the estimator of θ as $\hat{\theta}_{\text{snap}}(b)$ in the bth trial, $b \in [B]$.

For $\hat{\theta}_{\text{mle}}$, we calculate the bias, MSE and the percentage that θ falls in the CI(b) with $\alpha = 0.05$ in B trials; for $\hat{\theta}_{\text{snap}}$, we calculate the bias, MSE.

Table S1: Robostness of snapshot estimation.

Snapshot estimation		$p_w = \log(T)/T$		$p_w = \log^2(T)/T$	
		$\sum_{b=1}^{B} \hat{\theta}_{\text{snap}}(b)/B - \theta$	$\sum_{b=1}^{B} (\hat{\theta}_{\text{snap}}(b) - \theta)^2/B$	$\sum_{b=1}^{B} \hat{\theta}_{\text{snap}}(b)/B - \theta$	$\sum_{b=1}^{B} (\hat{\theta}_{\text{snap}}(b) - \theta)^2/B$
$\theta = 0.8$	$T = 10000$	$-9.0245e-04$	$2.3951e-05$	$2.8599e-03$	$3.8138e-05$
	$T = 2500$	$1.4657e-04$	$1.1062e-04$	$7.1498e-03$	$1.5758e-04$
	$T = 500$	$3.8589e-03$	$5.2520e-04$	0.0209	$1.1313e-03$
$\theta = 0.5$	$T = 10000$	$-1.7139e-04$	$2.9425e-05$	$7.0222e-03$	$9.1587e-05$
	$T = 2500$	$2.9211e-03$	$1.4328e-04$	0.0189	$5.2152e-04$
	$T = 500$	$3.4366e-03$	$6.5201e-04$	0.0514	$3.7241e-03$
$\theta = 0.2$	$T = 10000$	$1.5474e-03$	$2.2855e-05$	0.0108	$1.6576e-04$
	$T = 2500$	$4.2586e-03$	$1.3380e-04$	0.0296	$1.0971e-03$
	$T = 500$	0.0122	$6.1298e-04$	0.0824	$8.2613e-03$

Table S2: Robostness of MLE.

MLE		$p_w = \log(T)/T$			$p_w = \log^2(T)/T$		
		$\sum_{b=1}^{B} \hat{\theta}_{\text{mle}}(b)/B - \theta$	$\sum_{b=1}^{B} (\hat{\theta}_{\text{mle}}(b) - \theta)^2/B$	$\sum_{b=1}^{B} \mathbb{1}\{\theta \in \text{CI}(b)\}/B$	$\sum_{b=1}^{B} \hat{\theta}_{\text{mle}}(b)/B - \theta$	$\sum_{b=1}^{B} (\hat{\theta}_{\text{mle}}(b) - \theta)^2/B$	$\sum_{b=1}^{B} \mathbb{1}\{\theta \in \text{CI}(b)\}/B$
$\theta = 0.8$	$T = 10000$	$2.4131 - 04$	$2.8370e-05$	0.94	$3.1789e-03$	$3.6620e-05$	0.902
	$T = 2500$	$1.7276e-03$	$1.0523e-04$	0.956	$7.1773e-03$	$1.6139e-04$	0.886
	$T = 500$	$3.9507e-03$	$5.5523e-04$	0.948	0.0182	$8.7109e-04$	0.852
$\theta = 0.5$	$T = 10000$	$2.9466e-04$	$2.9230e-05$	0.956	$7.2660e-03$	$9.3111e-05$	0.734
	$T = 2500$	$3.0184e-03$	$1.3301e-04$	0.944	0.0180	$5.2736e-04$	0.61
	$T = 500$	$8.3624e-03$	$7.5881e-04$	0.912	0.0473	$3.2204e-03$	0.514
$\theta = 0.2$	$T = 10000$	$7.9673e-04$	$2.2896e-05$	0.934	0.0117	$1.9301e-04$	0.38
	$T = 2500$	$4.0064e-03$	$1.0821e-04$	0.904	0.0299	$1.1500e-03$	0.202
	$T = 500$	0.0128	$7.1726e-04$	0.868	0.0751	$6.8128e-03$	0.116

S3.2 Simulation 2 for Erroneous group labels

This subsection tests the estimators' robustness when the mislabeled nodes have high degrees. In order to explore this impact, set the nodes' observed group labels added in the time range $[1, t_0]$ of the evolving network are randomly sampled from $[K]$.

$$\breve{g}_i = \begin{cases} g_i, & \text{if node } i \text{ added at time } t \in [t_0+1, T]; \\ \text{a random sampling from } [K], & \text{if node } i \text{ added at time } t \in [1, t_0]. \end{cases}$$

Set $t_0 \in \{5, 15\}$, and others are same to the simulation in Section S3.1.

Table S3: Robostness of snapshot estimation.

		$t_0 = 5$		$t_0 = 15$	
		$\sum_{b=1}^{B} \hat{\theta}_{\text{snap}}(b)/B - \theta$	$\sum_{b=1}^{B} (\hat{\theta}_{\text{snap}}(b) - \theta)^2/B$	$\sum_{b=1}^{B} \hat{\theta}_{\text{snap}}(b)/B - \theta$	$\sum_{b=1}^{B} (\hat{\theta}_{\text{snap}}(b) - \theta)^2/B$
$\theta = 0.8$	$T = 10000$	$7.4914e-03$	$1.1540e-04$	0.0193	$4.5285e-04$
	$T = 2500$	0.0108	$2.6871e-04$	0.0250	$8.0679e-04$
	$T = 500$	0.0189	$1.0535e-03$	0.0359	$1.903e-03$
$\theta = 0.5$	$T = 10000$	0.0177	$5.0293e-04$	0.0456	$2.4332e-03$
	$T = 2500$	0.0244	$9.8794e-04$	0.0603	$4.340e-03$
	$T = 500$	0.0381	$2.6807e-03$	0.0909	$9.9104e-03$
$\theta = 0.2$	$T = 10000$	0.0272	$1.1729e-03$	0.0691	$5.4718e-03$
	$T = 2500$	0.0359	$1.9880e-03$	0.0946	0.0100
	$T = 500$	0.0561	$5.0639e-03$	0.1330	0.0205

Table S4: Robostness of MLE.

		$t_0 = 5$			$t_0 = 15$		
MLE		$\sum_{b=1}^{B} \hat{\theta}_{\mathrm{mle}}(b)/B - \theta$	$\sum_{b=1}^{B} (\hat{\theta}_{\mathrm{mle}}(b) - \theta)^2/B$	$\sum_{b=1}^{B} \mathbb{1}\{\theta \in \mathrm{CI}(b)\}/B$	$\sum_{b=1}^{B} \hat{\theta}_{\mathrm{mle}(b)}/B - \theta$	$\sum_{b=1}^{B} (\hat{\theta}_{\mathrm{mle}}(b) - \theta)^2/B$	$\sum_{b=1}^{B} \mathbb{1}\{\theta \in \mathrm{CI}(b)\}/B$
$\theta = 0.8$	$T = 10000$	$8.6303e-03$	$1.3336e-04$	0.592	0.0185	$4.2140e-04$	0.18
	$T = 2500$	0.0115	$2.8288e-04$	0.756	0.0264	$8.8387e-04$	0.308
	$T = 500$	0.0132	$7.6817e-04$	0.888	0.0373	$1.9426e-03$	0.602
$\theta = 0.5$	$T = 10000$	0.0193	$5.6158e-04$	0.3	0.0463	$2.5484e-03$	0.032
	$T = 2500$	0.0265	$1.1169e-03$	0.442	0.0625	$4.6289e-03$	0.064
	$T = 500$	0.0346	$2.4954e-03$	0.642	0.0923	0.0102	0.128
$\theta = 0.2$	$T = 10000$	0.0293	$1.2781e-03$	0.16	0.0661	$4.9474e-03$	0.0
	$T = 2500$	0.0381	$2.1829e-03$	0.25	0.0945	0.0102	0.01
	$T = 500$	0.0581	$5.0317e-03$	0.326	0.1357	0.0213	0.024

S3.3 Simulation for Missing of observation

We assume that the edge added to the network with a new node at vertex-step has a certain probability $p_{m,1}$ of being missed, and the edge appearing at edge-step has a certain probability $p_{m,2}$ of being missed.

For the evolving network $\{G(t) = (\mathcal{V}(t), \mathcal{E}(t), \mathcal{G}(t))\}_{t=0}^{T}$,

$$\{\tilde{G}(t') = (\tilde{\mathcal{V}}(t'), \tilde{\mathcal{E}}(t'), \mathcal{G}(t'))\}_{t'=0}^{T'}$$

is the corresponding observed evolving network. Moreover, let $\mathcal{M}$ be the set of nodes with missing observation. Missing observation will lead the length of the observed evolving network's time range (T') to be smaller than it is (T).

In detail, consider the procees of vertex-step and edge-step under missing of observation. Assume $\tilde{G}(0) = G(0)$, $t' = 0$, for $t = 1, 2, \cdots, T$:

- if $v(t) = 1$ (vertex-step), the new node w arrives with connected to an older node u:

$$\begin{cases} t' \text{ does not change}, & \text{if the edge } (w,u) \text{ is missed or node } u \in \mathcal{M}; \\ t' = t'+1, \tilde{\mathcal{V}}(t') = \tilde{\mathcal{V}}(t'-1) \cup w, & \text{else.} \end{cases}$$

$$\begin{cases} t' \text{ does not change}, & \text{if the edge } (w,u) \text{ is missed or node } u \in \mathcal{M}; \\ t' = t'+1, \tilde{\mathcal{E}}(t') = \tilde{\mathcal{E}}(t'-1) \cup (w,u), & \text{else.} \end{cases}$$

$$\begin{cases} \mathcal{M} = \mathcal{M} \cup w, & \text{if the edge } (w,u) \text{ is missed or node } u \in \mathcal{M}; \\ \mathcal{M} \text{ does not change}, & \text{else.} \end{cases}$$

- if $v(t) = 0$ (edge-step), the new edge (w,u) occurs between older nodes w, u:

$$\begin{cases} t' \text{ does not change}, & \text{if node } w \in \mathcal{M} \text{ or } (w,u) \text{ is missed;} \\ t' = t'+1, \tilde{\mathcal{V}}(t') = \tilde{\mathcal{V}}(t'-1), & \text{else if } u \notin \mathcal{M}; \\ t' = t'+1, \tilde{\mathcal{V}}(t') = \tilde{\mathcal{V}}(t'-1) \cup u, & \text{else.} \end{cases}$$

$$\begin{cases} t' \text{ does not change}, & \text{if nodes } w \in \mathcal{M} \text{ or } (w,u) \text{ is missed;} \\ t' = t'+1, \tilde{\mathcal{E}}(t') = \tilde{\mathcal{E}}(t'-1) \cup (w,u), & \text{else.} \end{cases}$$

$$\begin{cases} \mathcal{M} = \mathcal{M}/u, & \text{if } w \notin \mathcal{M}, u \in \mathcal{M} \text{ and } (w,u) \text{ is not missed;} \\ \mathcal{M} \text{ does not change,} & \text{else.} \end{cases}$$

Run the simulation in B trials, $B = 500$.

- Set $T \in \{500, 1000, 2500\}$. The number of groups $K = 10$. The probability of missing $p_{w,1} = p_{w,2} \in \{0.1, 0.5\}$.
- The initial graph has 10 nodes, and $10 \times p_k$ nodes are from group k, $k \in [K]$, where $p = (\underbrace{0.1, 0.1 \cdots}_{10})$.
- For each time t, a vertex-step arrives with probability q, $q = 0.5$.
- The procees of vertex-step and edge-step under missing of observation is as described above. Record the $\{v(t'),\, e(t') = (e_1(t'), e_2(t'))\}$ at each time t' if t' has changed.
- $\{D_k(t)\}_{k=1}^K$ are the group degrees with graph $G(t)$ used to generate the evolving network; $\{\tilde{D}_k(t')\}_{k=1}^K$ are the observed group degrees with $\tilde{\mathcal{V}}(t'), \tilde{\mathcal{E}}(t')$.
- $\{\{\tilde{D}_k(t')\}_{k=1}^K, v(t'), g_{e_1(t')}, g_{e_2(t')}\}_{t'=0}^{T'}$ are used to construct the maximum likelihood estimation (Equation (4.1) in the main paper). Record the estimator of θ as $\hat{\theta}_{\text{mle}}(b)$ in the bth trial, $b \in [B]$.

- For the graph $\tilde{G}(T')$, $\tilde{\mathcal{E}}_{k,0}(T')$ is the set of edges from group k to another, and $\tilde{\mathcal{E}}_{k,1}(T')$ is the set of edges that both nodes from group k under missing of observation.

- $\{\tilde{D}_k(T'), \tilde{\mathcal{E}}_{k,1}(T'), \tilde{\mathcal{E}}_{k,0}(T')\}_{k=1}^{K}$ are used to construct the snapshot estimation (Equation (4.9) in the main paper). Record the estimator of θ as $\hat{\theta}_{\text{snap}}(b)$ in the bth trial, $b \in [B]$.

- Record the bias $\sum_{b=1}^{B}(\hat{\theta}(b) - \theta)/B$, MSE $\sum_{b=1}^{B}(\hat{\theta}(b) - \theta)^2/B$, and Cover rate $\sum_{b=1}^{B} \mathbb{1}\{\theta \in \text{CI}(b)\}/B$ of the estimator $\hat{\theta}$ with mislabeling in B trials, where $\text{CI}(b)$ is the 95% confidence interval of $\hat{\theta}(b)$.

Table S5: Robustness of MLE.

		$p_{m,1} = p_{m,2} = 0.1$			$p_{m,1} = p_{m,2} = 0.5$		
		$\sum_{b=1}^{B} \hat{\theta}_{\text{mle}}(b)/B - \theta$	$\sum_{b=1}^{B}(\hat{\theta}_{\text{mle}}(b) - \theta)^2/B$	$\sum_{b=1}^{B} \mathbb{1}\{\theta \in \text{CI}(b)\}/B$	$\sum_{b=1}^{B} \hat{\theta}_{\text{mle}}(b)/B - \theta$	$\sum_{b=1}^{B}(\hat{\theta}_{\text{mle}}(b) - \theta)^2/B$	$\sum_{b=1}^{B} \mathbb{1}\{\theta \in \text{CI}(b)\}/B$
$\theta = 0.8$	$T = 2500$	$1.0723e - 03$	$1.2708e - 04$	0.942	$3.8026e - 03$	$2.5154e - 04$	0.934
	$T = 1000$	$3.0262e - 04$	$2.9416e - 04$	0.956	$4.3161e - 03$	$6.9197e - 04$	0.936
	$T = 500$	$6.3454e - 04$	$5.6375e - 04$	0.952	$4.4283e - 03$	$1.2044e - 03$	0.952
$\theta = 0.5$	$T = 2500$	$-1.0565e - 05$	$1.3795e - 04$	0.962	$3.8667e - 03$	$2.6979e - 04$	0.962
	$T = 1000$	$-3.3123e - 05$	$3.4095e - 04$	0.96	$3.9199e - 03$	$7.4547e - 04$	0.956
	$T = 500$	$-1.7034e - 04$	$7.6153e - 04$	0.946	$5.4110e - 03$	$1.4387e - 03$	0.952
$\theta = 0.2$	$T = 2500$	$-4.1849e - 04$	$8.5990e - 05$	0.948	$2.0122e - 03$	$1.9869e - 04$	0.936
	$T = 1000$	$9.8879e - 04$	$1.9708e - 04$	0.966	$9.3412e - 04$	$4.9876e - 04$	0.956
	$T = 500$	$-1.0030e - 04$	$4.5918e - 04$	0.94	$3.1249e - 03$	$1.0362e - 03$	0.944

Table S6: Robustness of snapshot estimation.

		$p_{m,1} = p_{m,2} = 0.1$		$p_{m,1} = p_{m,2} = 0.5$	
		$\sum_{b=1}^{B} \hat{\theta}_{\text{snap}}(b)/B - \theta$	$\sum_{b=1}^{B} (\hat{\theta}_{\text{snap}}(b) - \theta)^2/B$	$\sum_{b=1}^{B} \hat{\theta}_{\text{snap}}(b)/B - \theta$	$\sum_{b=1}^{B} (\hat{\theta}_{\text{snap}}(b) - \theta)^2/B$
$\theta = 0.8$	$T = 2500$	$-1.3145e-03$	$6.0570e-04$	$1.7764e-03$	$1.1082e-03$
	$T = 1000$	$-3.1679e-03$	$1.2020e-03$	$1.0909e-03$	$1.9871e-03$
	$T = 500$	$-4.0074e-03$	$1.8869e-03$	$5.8095e-03$	$3.2324e-03$
$\theta = 0.5$	$T = 2500$	$-1.7368e-03$	$2.7252e-04$	$1.9237e-03$	$5.4182e-04$
	$T = 1000$	$-2.1823e-03$	$5.5534e-04$	$2.9046e-03$	$1.1986e-03$
	$T = 500$	$-2.8367e-03$	$1.1384e-03$	$3.6887e-03$	$2.2657e-03$
$\theta = 0.2$	$T = 2500$	$-8.3171e-04$	$1.0330e-04$	$1.3001e-03$	$2.3001e-04$
	$T = 1000$	$7.9986e-05$	$2.3972e-04$	$-2.6247e-04$	$5.5676e-04$
	$T = 500$	$-1.01286e-03$	$5.0915e-04$	$2.8002e-03$	$1.1757e-03$

S3.4 Group label recovery

This subsection is to verify Theorem S.2. We use the method in Section S2 to recover the labels of the 10 nodes with the largest degrees. Set $B = 500$, $T \in \{500, 2500, 10000\}$, $K \in \{3, 5, 10\}$, $n_0 = 100$, $q = 0.5$, $p = p(K)$ in Equation (7.1) in the main paper.

$\mathcal{M}_T$ is the 10 nodes with the largest degrees in graph $G(T)$ in each trial. Estimate the group labels of nodes in $\mathcal{M}_T$ by Equation (S2.1). "Rate" records the percentage of successful recoveries in B trials — $\sum_{i \in \mathcal{M}_T} \sum_{b=1}^{B} \mathbb{1}\{\hat{g}_i(b) = g_i(b)\}/(10B)$; "Childset" records the average of the size of 10 largest degrees nodes' childsets in B trials — $\sum_{i\in\mathcal{M}_T}\sum_{b=1}^{B}|\mathcal{C}_i(b)|/(10B)$. The results are recorded in Table S7.

Table S7: The performance of group label recovery.

		Rate			Childset		
		$\theta=0.9$	$\theta=0.7$	$\theta=0.5$	$\theta=0.9$	$\theta=0.7$	$\theta=0.5$
$K=3$	$T=10000$	0.6407	0.9667	1	58.92	58.3687	58.4613
	$T=2500$	0.542	0.836	0.9673	20.7227	20.352	20.1067
	$T=500$	0.44	0.6367	0.8007	5.9473	5.8553	5.9667
$K=5$	$T=10000$	0.5513	0.9293	0.9933	59.266	58.4847	58.4593
	$T=2500$	0.3873	0.754	0.9393	20.452	20.0467	20.4126
	$T=500$	0.2667	0.52	0.7087	5.9273	5.8907	5.8207
$K=10$	$T=10000$	0.504	0.946	0.9987	58.54	57.87	56.9087
	$T=2500$	0.298	0.7593	0.95	20.3307	20.426	20.142
	$T=500$	0.178	0.4307	0.6807	6.0467	5.7133	5.8687

S3.5 Simulation for Instability

Set $B=500$, $K=10$, $n_0=100$, $p=(0.2,0.2,0.1,0.1,0.1,\underbrace{0.06,0.06,\cdots}_{5})$.

$$q=\begin{cases} q_1, & t<T/3 \\ q_2, & T/3\le t<2T/3 \\ q_3, & t\ge 2T/3. \end{cases}$$

$\{\{D_k(t)\}_{k=1}^{K}, v(t), g_{e_1(t)}, g_{e_2(t)}\}_{t=1}^{T}$ are used to construct the maximum likelihood equation (Equation (4.1) in the main paper) and record the estimator of θ as $\hat{\theta}_{\text{mle}}(b)$ in the bth trial, $b \in [B]$. $\{D_k(T), \mathcal{E}_{k,1}(T), \mathcal{E}_{k,0}(T)\}_{k=1}^K$ are used to construct the snapshot estimation (Equation (4.9) in the main paper) and record the estimator of θ as $\hat{\theta}_{\text{snap}}(b)$ in bth trial, $b \in [B]$.

"Bias" records the absolute sum of the bias from B trials: $\sum_{k=1}^K \left| \sum_{b=1}^B [D_{k,b}(T)/(2T) - p_k] \right| /B$. "MSE" records the sum of the mean square error from B trials: $\sum_{k=1}^K \sum_{b=1}^B [D_{k,b}(T)/(2T) - p_k]^2/B$.

Table S8: Behavior of $D_k(T)/2T$ with change points (q_1, q_2, q_3).

		Bias			MSE		
	(q_1,q_2,q_3)	$(0.9,0.5,0.1)$	$(0.9,0.9,0.5)$	$(0.9,0.9,0.9)$	$(0.9,0.5,0.1)$	$(0.9,0.9,0.5)$	(0.9,0.9,0.9)
$T=1000$	$\theta=0.8$	$4.5105e-03$	$5.4971e-03$	$1.8514e-03$	$2.3320e-03$	$1.5562e-03$	$1.3299e-03$
	$\theta=0.5$	$6.0286e-03$	$7.2495e-03$	$3.2419e-03$	$2.3062e-03$	$1.4550e-03$	$1.2021e-03$
	$\theta=0.2$	$6.5962e-03$	$6.2324e-03$	$1.8171e-03$	$2.2928e-03$	$1.3628e-03$	$9.8406e-04$
$T=5000$	$\theta=0.8$	$2.4642e-03$	$2.8602e-03$	$2.1057e-03$	$7.1793e-04$	$4.9953e-04$	$3.9480e-04$
	$\theta=0.5$	$2.3525e-03$	$1.0681e-03$	$1.8448e-03$	$6.2593e-04$	$3.5772e-04$	$2.8718e-04$
	$\theta=0.2$	$3.2840e-03$	$1.8392e-03$	$1.9299e-03$	$5.6204e-04$	$3.0870e-04$	$2.3889e-04$
$T=25000$	$\theta=0.8$	$9.5952e-04$	$1.4892e-03$	$5.5569e-04$	$1.9955e-04$	$1.2470e-04$	$1.0048e-04$
	$\theta=0.5$	$1.4724e-03$	$9.3613e-04$	$1.6771e-03$	$1.4184e-04$	$8.1772e-05$	$6.8857e-05$
	$\theta=0.2$	$9.1705e-04$	$8.5277e-04$	$7.7605e-04$	$1.1649e-04$	$6.7630e-05$	$4.9284e-05$

Fix $(q_1, q_2, q_3) = (0.9, 0.5, 0.1)$. We calculate the mean absolute error (MAE) and the mean square error (MSE) for $\hat{\theta}_{\text{mle}} = \sum_{b=1}^B \hat{\theta}_{\text{mle}}(b)/B$ and $\hat{\theta}_{\text{snap}} = \sum_{b=1}^B \hat{\theta}_{\text{snap}}(b)/B$, $b \in [B]$.

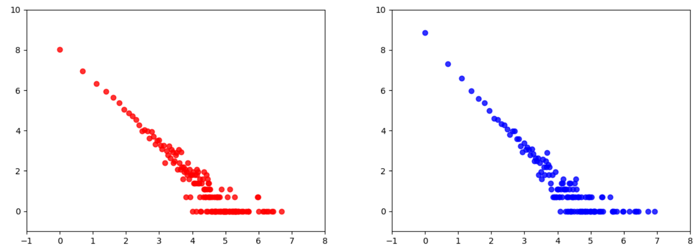
Table S9: Comparison of the MLE and snapshot estimations.

		Bias		MAE		MSE	
		$\hat{\theta}_{\text{mle}} - \theta$	$\hat{\theta}_{\text{snap}} - \theta$	$\sum_{b=1}^{B} \lvert\hat{\theta}_{\text{snap}}(b) - \theta\rvert/B$	$\sum_{b=1}^{B} \lvert\hat{\theta}_{\text{mle}}(b) - \theta\rvert/B$	$\sum_{b=1}^{B}(\hat{\theta}_{\text{snap}}(b) - \theta)^2/B$	$\sum_{b=1}^{B}(\hat{\theta}_{\text{mle}}(b) - \theta)^2/B$
$T = 1000$	$\theta = 0.8$	$1.2415e-03$	$2.6980e-03$	0.0131	0.0133	$2.7294e-04$	$2.8096e-04$
	$\theta = 0.5$	$5.6497e-04$	$1.3591e-03$	0.0145	0.0145	$3.3325e-04$	$3.3676e-04$
	$\theta = 0.2$	$1.2478e-04$	$4.0973e-04$	0.0113	0.0113	$1.9960e-04$	$2.0033e-04$
$T = 5000$	$\theta = 0.8$	$5.7625e-04$	$8.6632e-04$	$6.0536e-03$	$6.1008e-03$	$5.7673e-05$	$5.8839e-05$
	$\theta = 0.5$	$3.4268e-04$	$5.4127e-4$	$6.4096e-03$	$6.4628e-03$	$6.4362e-05$	$6.4908e-05$
	$\theta = 0.2$	$-1.0038e-04$	$-3.2085e-05$	$4.9376e-03$	$4.9393e-03$	$3.6573e-05$	$3.6620e-05$

S4. Data application

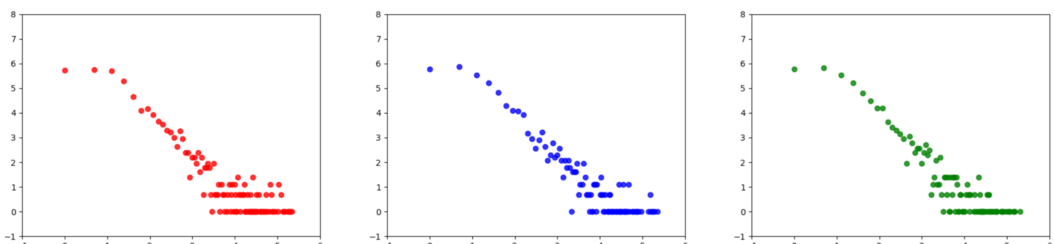
Table S10: Information about network datasets.

Name	CL-10K-1d8-L5	soc-political-retweet
Number of nodes	10000	18470
Number of groups	5	2
Number of nodes from group 1	2000	7115
Number of nodes from group 2	2000	11355
Number of nodes from group 3	2000	\
Number of nodes from group 4	2000	\
Number of nodes from group 5	2000	\
Number of edges	44896	61157
Total degrees from group 1	18965	60139
Total degrees from group 2	17381	49187
Total degrees from group 3	17309	\
Total degrees from group 4	17949	\
Total degrees from group 5	18188	\
Estimation method	snapshot	MLE
The estimator $\hat{\theta}$	0.9999	0.0413

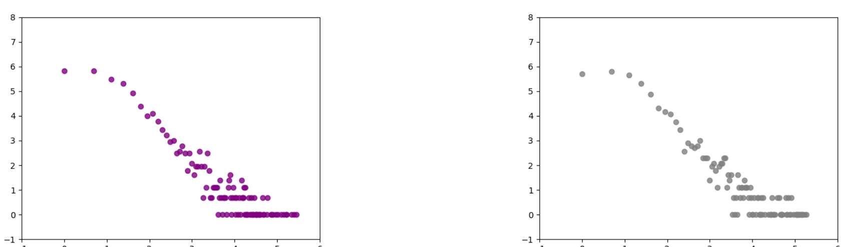


(a) Group 1 (b) Group 2

Figure S2: Log-log degree distribution from different groups of soc-political-retweet.



(a) Group 1 (b) Group 2 (c) Group 3



(d) Group 4 (e) Group 5

Figure S3: Log-log degree distribution from different groups of CL-10K-1d8-L5.

S5. Proofs.

Proof of Theorem 1. Let $Z_k(t) = D_k(t) - D_k(t-1), t \geq 1$ be the increased degree of group k at time t. We define the $\sigma-$algebra $\mathcal{F}_t = \sigma(G(0), \cdots, G(t))$. Let $P_k(t) = \frac{D_k(t-1)}{2(t-1)+n_0}$, we have

$$\begin{aligned} E[Z_k(t)|\mathcal{F}_{t-1}] = \quad & 2qp_k\left[P_k(t) + (1-\theta)\left(1-P_k(t)\right)\right] + \theta q(1-p_k)P_k(t) + \theta qp_k(1-P_k(t)) \\ & +2(1-q)P_k(t)\left[P_k(t) + (1-\theta)\left(1-P_k(t)\right)\right] + 2\theta(1-q)(1-P_k(t))P_k(t) \\ = \quad & 2qp_k - qp_k\theta\left(1-P_k(t)\right) + [\theta q(1-p_k) + 2(1-q)]P_k(t) \\ = \quad & (2-\theta)qp_k + [2-q(2-\theta)]P_k(t). \end{aligned} \tag{S5.2}$$

Let $Z_k^0(i) := Z_k(i) - E[Z_k(i)|\mathcal{F}_{i-1}]$, $S_k(t) := \sum_{i=1}^t Z_k^0(i)$. By Chebyshev's inequality, for any $\epsilon > 0$,

$$P\left(\left|\frac{S_k(t^2)}{t^2}\right| > \epsilon\right) \leq \frac{E[(S_k(t^2))^2]}{t^4\epsilon^2} = \frac{\sum_{i=1}^{t^2} E[(Z_k^0(i))^2] + \sum_{0<i\neq j\leq t} E[Z_k^0(i)Z_k^0(j)]}{t^4\epsilon^2}. \tag{S5.3}$$

Further, $E[(Z_k^0(t))^2] \leq 4$ and for any $i > j$,

$$\begin{aligned} & E[Z_k^0(i)Z_k^0(j)] = E\left[E[Z_k^0(i)Z_k^0(j)|\mathcal{F}_{i-1}]\right] \\ = \quad & E\left[Z_k^0(j)E[Z_k^0(i)|\mathcal{F}_{i-1}]\right] = 0. \end{aligned} \tag{S5.4}$$

By Equations (S5.3)–(S5.4), we can conclude that

$$P\left(\left|\frac{S_k(t^2)}{t^2}\right| > \epsilon\right) \leq \frac{4t^2}{t^4\epsilon} \leq \frac{c}{t^2}. \tag{S5.5}$$

Equation (S5.5) implies $\sum_{t=1}^{\infty} P\left(\left|\frac{S_k(t^2)}{t^2}\right| > \epsilon\right) < \infty$. Further, we can get

$$\frac{S_k(t^2)}{t^2} \xrightarrow{a.s.} 0. \tag{S5.6}$$

From Equation (S5.6), $\forall t \geq 1$, there exists m such that

$$\begin{aligned}
& \max_{m^2 \leq t < (m+1)^2} \left|\frac{S_k(t)}{t}\right| \leq \frac{1}{m^2} \max_{m^2 \leq t < (m+1)^2} |S_k(t)| \\
\leq \quad & \frac{1}{m^2}\left(|S_k(m^2)| + \max_{m^2 \leq t < (m+1)^2} |S_k(t) - S_k(m^2)|\right) \\
\leq \quad & \frac{1}{m^2}\left(|S_k(m^2)| + 2((m+1)^2 - m^2)\right) \\
\leq \quad & \frac{|S_k(m^2)|}{m^2} + \frac{4m+2}{m^2}.
\end{aligned} \tag{S5.7}$$

Equation (S5.7) implies $\frac{S_k(t)}{t} \xrightarrow{a.s.} 0$. Next, by the convergence theorem,

$$\begin{aligned}
\lim_{t\to\infty} \frac{D_t^k}{2t} = \quad & \lim_{t\to\infty} \left(\frac{n_0 p_k + \sum_{i=1}^t Z_i^k}{2t}\right) \\
= \quad & \lim_{t\to\infty} \left[\frac{n_0 p_k + S_k(t) + \sum_{i=1}^t E(Z_i^k|\mathcal{F}_{i-1})}{2t}\right] \\
= \quad & \lim_{t\to\infty} \frac{\sum_{i=1}^t E(Z_i^k|\mathcal{F}_{i-1})}{2t}.
\end{aligned} \tag{S5.8}$$

When t is large enough, we can conclude from Equations (S5.2) and (S5.8) that

$$\begin{aligned}
\frac{D_k(t)}{2t+n_0} = \quad & \frac{\sum_{i=1}^t E(Z_k(i)|\mathcal{F}_{i-1})}{2t} + o(1) \\
= \quad & \frac{(2-\theta)qp_k}{2} + \frac{[2-q(2-\theta)]}{2t} \sum_{i=1}^t \frac{D_k(i-1)}{2(i-1)+n_0} + o(1).
\end{aligned} \tag{S5.9}$$

According to the Equation (S5.9), let:

$$f_k(x) = \frac{(2-\theta)qp_k}{2} + \frac{[2-q(2-\theta)]}{2}x. \tag{S5.10}$$

By the Banach fixed point theorem,

$f_k(x): (R, |\cdot|) \to (R, |\cdot|)$ is a contraction mapping with only one fixed point $x = p_k$.

By Equation (S5.10), when t is large enough, we have:

$$\left|\frac{D_k(t)}{2t+n_0} - p_k\right| = \frac{[2-q(2-\theta)]}{2}\left|\frac{\sum_{i=1}^{t}\left[\frac{D_k(i-1)}{2(i-1)+n_0} - p_k\right]}{t}\right| + o(1), \tag{S5.11}$$

where $0 < |2-q(2-\theta)|/2 < 1$.

Equation (S5.11) implies that $\frac{D_k(t)}{2t+n_0}$ approaches p_k as $t \to \infty$. By Equations (S5.9)–(S5.11), we can deduce $\frac{D_k(t)}{2t} \xrightarrow{a.s.} p_k$. □

Proof of Corollary 1. For any node $i \in G(t_0)$, $g_i = k$, let $Z_i(t+1) = d_i(t+$

$1) - d_i(t)$, $P_i(t) := \frac{d_i(t)}{2t+n_0}$ for $t \in [t_0+1, T]$.

$$\begin{aligned}
E[Z_i(t+1)|\mathcal{F}_t] = \quad & q\left[p_k\theta P_i(t) + p_k(1-\theta)P_i(t)\frac{2t+n_0}{D_k(t)}\right] + q(1-p_k)\theta P_i(t) \\
& +(1-q)\theta P_i(t)\frac{D_k(t)}{2t+n_0} + \theta(P_i(t))^2 + (1-\theta)P_i(t) + (1-\theta)(P_i(t))^2\frac{2t+n_0}{D_k(t)} \\
& +(1-q)\left[\theta P_i(t) - \theta P_i(t)\frac{D_k(t)}{2t+n_0}\right] \\
& +(1-q)\left[\theta P_i(t)\frac{D_k(t)}{2t+n_0} - \theta(P_i(t))^2 + (1-\theta)P_i(t) - (1-\theta)(P_i(t))^2\frac{2t+n_0}{D_k(t)}\right] \\
& +(1-q)\left[\theta P_i(t) - \theta P_i(t)\frac{D_k(t)}{2t+n_0}\right] \\
= \quad & q\theta P_i(t) + qp_k(1-\theta)P_i(t)\frac{2t+n_0}{D_k(t)} + 2(1-q)P_i(t) \\
= \quad & (2-q)P_i(t) + \left[1 - p_k\frac{2t+n_0}{D_k(t)}\right]q(1-\theta)P_i(t).
\end{aligned}$$

Let $Z_i^0(t) := Z_i(t) - E[Z_i(t)|\mathcal{F}_{t-1}]$, $E[(Z_i^0(t))^2] \leq 1$ and $S_i^0(t) := \sum_{j=t_0+1}^{t} Z_i^0(j)$.

By the steps of Equations (S5.3)-(S5.6), we get $S_i^0(t)/t \xrightarrow{a.s.} 0$ and

$$\lim_{t\to\infty}\frac{d_i(t)}{2t} = \lim_{t\to\infty}\frac{d_i(t_0) + \sum_{j=t_0+1}^{t} Z_i^0(j)}{2t} = \lim_{t\to\infty}\frac{\sum_{j=t_0+1}^{t} E[Z_i(j)|\mathcal{F}_{j-1}]}{2t}.$$

Further, we can get:

$$\frac{d_i(t)}{2t+n_0} = \frac{\sum_{j=t_0+1}^{t-1}\left[(2-q) + (1-p_k\frac{2j+n_0}{D_k(j)})q(1-\theta)\right]\frac{d_i(j)}{2j+n_0}}{2t+n_0} + o(1).$$

By Theorem 1, we have $\frac{D_k(t)}{2t+n_0} \xrightarrow{a.s.} p_k$. It implies that $\left|\frac{D_t^k}{2t+n_0} - p_k\right| \geq \epsilon$ with probability $o_p(1)$ for any $\epsilon > 0$. Set $\epsilon = \frac{p_k}{2(1-\theta)}$, then

$$\left|\frac{d_i(t)}{2t+n_0}\right| \leq (1-q/4)\left|\frac{\sum_{j=t_0+1}^{t}\frac{d_i(j)}{2j+n_0}}{t-t_0}\right| + o(1), \tag{S5.12}$$

with probability $1 - o_p(1)$.

Equation (S5.12) implies $\frac{d_i(t)}{2t} \xrightarrow{i.p.} 0$ for any fixed node i. □

Proof of Corollary 2. By the same operations as in Equations (S5.2)–(S5.7), we have:

$$\frac{\sum_{j=1}^{t}[X(j) - E(X(j)|\mathcal{F}_{j-1})]}{t} \xrightarrow{a.s.} 0.$$

Then:

$$\frac{S(t)}{t} = \frac{\sum_{j=1}^{t}[X(j) - E(X(j)|\mathcal{F}_{j-1})] + \sum_{j=1}^{t} E(X(j)|\mathcal{F}_{j-1})}{t} \xrightarrow{a.s.} \lim_{t\to\infty} \frac{\sum_{j=1}^{t} E(X(j)|\mathcal{F}_{j-1})}{t}.$$

Let $P_k(t) := \frac{D_k(t-1)}{2(t-1)+n_0}$, we can get:

$$\begin{aligned} & E(X(t)|\mathcal{F}_{t-1}) \\ = \quad & q\sum_{k=1}^{K} p_k \left[P_k(t) + (1-\theta)\left(1-P_k(t)\right)\right] + (1-q)\sum_{k=1}^{K} P_k(t)\left[P_k(t) + (1-\theta)\left(1-P_k(t)\right)\right]. \end{aligned}$$

By Theorem 1, we have $\frac{D_k(t)}{2t} \xrightarrow{a.s.} p_k$, so

$$\begin{aligned} \frac{S(t)}{t} \xrightarrow{a.s.} \lim_{t\to\infty} \frac{\sum_{i=1}^{t} E(X(i)|\mathcal{F}_{i-1})}{t} &= \lim_{t\to\infty} E(X(t)|\mathcal{F}_{t-1}) \\ &= \sum_{k=1}^{K} p_k\left[1-\theta(1-p_k)\right] \\ &= 1+\theta\left(\sum_{k=1}^{K} p_k^2 - 1\right). \end{aligned}$$

□

Proof of Theorem 2.

$$\begin{aligned} E\left[D_k(t) - p_k(2t+n_0)\middle|\, \mathcal{F}_{t-1}\right] = & \left[1 + \frac{2-q(2-\theta)}{2(t-1)+n_0}\right] D_k(t-1) + (2-\theta)qp_k - p_k(2t+n_0) \\ = & \left[1 + \frac{2-q(2-\theta)}{2(t-1)+n_0}\right]\left[D_k(t-1) - p_k(2(t-1)+n_0)\right]. \end{aligned}$$

Let $X_k^0(t) = \frac{D_k(t)-p_k(2t+n_0)}{\prod_{i=1}^t[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}]}, \forall t \geq 1$. $X_k^0(0) = D_k(0) - p_k n_0 = 0$. Obviously, $\{X_k^0(t)\}_t$ is a martingale sequence.

$$
\begin{aligned}
X_k^0(t) - X_k^0(t-1) = \quad & \frac{D_k(t) - p_k(2t+n_0)}{\prod_{i=1}^t[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}]} - \frac{D_k(t-1) - p_k(2t-2+n_0)}{\prod_{i=1}^{t-1}[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}]} \\
= \quad & \frac{D_k(t) - p_k(2t+n_0) - [1+\frac{2-q(2-\theta)}{2(t-1)+n_0}][D_k(t-1) - p_k(2t-2+n_0)]}{\prod_{i=1}^t[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}]} \\
= \quad & \frac{[D_k(t) - D_k(t-1)] - 2p_k + \left[\frac{2-q(2-\theta)}{2(t-1)+n_0}\right][D_k(t-1) - p_k(2t-2+n_0))]}{\prod_{i=1}^t\left[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}\right]}.
\end{aligned}
$$

$D_k(t)$ is the total of the degrees from group k in the graph $G(t)$, so $0 \leq D_k(t) - D_k(t-1) \leq 2$, $1 \leq D_k(t-1) \leq 2t - 2 + n_0$. Then we have

$$
|X_k^0(t) - X_k^0(t-1)| \leq \frac{4}{\prod_{i=1}^t\left[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}\right]}.
$$

What is more, we get the following upper bound for $Var(X_k^0(t)|\mathcal{F}_{t-1})$:

$$
\begin{aligned}
Var(X_k^0(t)|\mathcal{F}_{t-1}) = \quad & \frac{Var(D_k(t) - p_k(2t+n_0)|\mathcal{F}_{t-1})}{\prod_{i=1}^t\left[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}\right]^2} = \frac{Var(D_k(t)|\mathcal{F}_{t-1})}{\prod_{i=1}^t\left[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}\right]^2} \\
= \quad & \frac{E[(D_k(t) - E(D_k(t)|\mathcal{F}_{t-1}))^2|\mathcal{F}_{t-1}]}{\prod_{i=1}^t\left[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}\right]^2} \\
\leq \quad & \frac{E[(D_k(t) - D_k(t-1))^2|\mathcal{F}_{t-1}]}{\prod_{i=1}^t\left[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}\right]^2} \\
\leq \quad & \frac{4}{\prod_{i=1}^t\left[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}\right]^2}.
\end{aligned}
$$

By Theorems 2.22 and 2.26 in Chung and Lu (2006), let $M_j = 4/\prod_{i=1}^j\left[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}\right]$

and $\sigma_j^2 = 4/\prod_{i=1}^{j}\left[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}\right]^2$. Then we get

$$P(|X_k^0(t) - EX_k^0(t)| \geq \lambda) \leq e^{-\frac{\lambda^2}{2\sum_{j=1}^{t}(\sigma_j^2+M_j^2)}}.$$

Note that $EX_k^0(t) = EX_k^0(1) = X_k^0(0) = 0$, so we have:

$$P(|X_k^0(t)| \geq \lambda) \leq e^{-\frac{\lambda^2}{2\sum_{j=1}^{t}(\sigma_j^2+M_j^2)}}. \tag{S5.13}$$

By Stirling's approximation,

$$\begin{aligned}\prod_{i=1}^{t}\left[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}\right] &= \prod_{i=1}^{t}\left[\frac{2(i-1)+n_0+2-q(2-\theta)}{2(i-1)+n_0}\right] \\ &= \frac{\prod_{i=1}^{t}[i+(n_0-q(2-\theta))/2]}{\prod_{i=1}^{t}[i+(n_0-2)/2]} \\ &= O(t^{1-q(2-\theta)/2}).\end{aligned} \tag{S5.14}$$

$$\prod_{i=1}^{t}\left[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}\right]^2 = O(t^{2-q(2-\theta)}). \tag{S5.15}$$

For $t \geq 1$, we have:

$$\sum_{j=1}^{t}(\sigma_j^2+M_j^2) = O\left(\sum_{j=1}^{t} j^{q(2-\theta)-2}\right). \tag{S5.16}$$

For $q(2-\theta)-2<0$, the function $f(x) = x^{q(2-\theta)-2}$ is strictly monotonically decreasing with respect to x on $(0, +\infty)$.

Then we have the inequality:

$$\int_1^t j^{q(2-\theta)-2}dj < \sum_{j=1}^{t} j^{q(2-\theta)-2} < 1 + \int_1^{t-1} j^{q(2-\theta)-2}dj. \tag{S5.17}$$

Further,

$$\int_1^t j^{q(2-\theta)-2} dj = \begin{cases} t^{q(2-\theta)-1} - 1, & \text{if } 1/(2-\theta) < q \le 1; \\ \log(t), & \text{if } q = 1/(2-\theta); \\ 1 - t^{q(2-\theta)-1}, & \text{if } 0 < q < 1/(2-\theta). \end{cases} \tag{S5.18}$$

Equations (S5.13)–(S5.18) imply that when t is large enough, we have

$$\sum_{j=1}^{t} j^{q(2-\theta)-2} = \begin{cases} O(t^{q(2-\theta)-1}), & \text{if } 1/(2-\theta) < q \le 1; \\ O(\log t), & \text{if } q = 1/(2-\theta); \\ O(1), & \text{if } 0 < q < 1/(2-\theta). \end{cases} \tag{S5.19}$$

Combine Equations (S5.13)–(S5.19), we can get:

(1) If $1/(2-\theta) < q \le 1$, let $\lambda = \frac{2c_1(t)t^{1/2}}{\prod_{i=1}^t \left[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}\right]}$, then

$$P\left(|X_k^0(t)| \ge \frac{2c_1(t)t^{1/2}}{\prod_{i=1}^t \left[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}\right]}\right) \le e^{-\frac{\lambda^2}{2\sum_{i=1}^t(\sigma_i^2+M_i^2)}} \propto e^{-C_1(c_1(t))^2}.$$

$$P\left(|D_k(t) - p_k(2t+n_0)| \ge 2c_1(t)t^{1/2}\right) \le e^{-C_1(c_1(t))^2}.$$

(2) Otherwise, if $q = 1/(2-\theta)$, let $\lambda = \frac{2c_2(t)(\log(t))^{1/2}}{\prod_{i=1}^t \left[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}\right]}$, then

$$P\left(|D_k(t) - p_k(2t+n_0)| \ge 2c_2(t)(\log(t))^{1/2}\right) \le e^{-C_2(c(t))^2/t}.$$

(3) Otherwise, if $0 < q < 1/(2-\theta)$, let $\lambda = \frac{2c_3(t)}{\prod_{i=1}^t \left[1+\frac{2-q(2-\theta)}{2(i-1)+n_0}\right]}$, then

$$P\left(|D_k(t) - p_k(2t+n_0)| \ge 2c_3(t)\right) \le e^{-C_3(c(t))^2 t^{q(2-\theta)-2}}.$$

C_1, C_2, C_3 are constants larger than 0, and $c_1(t), c_2(t), c_3(t)$ is a strictly monotonically increasing positive function of t. □

Lemma 1. *Suppose that there are stochastic processes $\{a_t\}$, $\{b_t\}$, and constant sequence $\{c_t\}$ satisfying the recurrence relation*

$$E(a_{t+1}) = E\left[\left(1 - \frac{b_t}{t+t_1}\right) a_t\right] + c_t,\ \forall t \geq t_0.$$

where $\lim_{t\to\infty} c_t = c$, $\lim_{t\to\infty} b_t \xrightarrow{a.s.} b > 0$ *and* $0 < a_t/t \leq C$, C *is a positive constant. Then* $\lim_{t\to\infty} E(a_t)/t$ *exists and*

$$\lim_{t\to\infty} \frac{E(a_t)}{t} = \frac{c}{1+b}.$$

Proof of Lemma 1. Without loss of generality, we can assume $t_1 = 0$ after shifting t by t_1. By rearranging the recurrence relation, we have

$$\begin{aligned} \frac{E(a_{t+1})}{t+1} - \frac{c}{1+b} &= \frac{E\left[\left(1 - \frac{b_t}{t}\right) a_t\right] + c_t}{t+1} - \frac{c}{1+b} \\ &= \left[\frac{E(a_t)}{t} - \frac{c}{1+b}\right]\left(1 - \frac{1+b}{t+1}\right) + \frac{c_t - c}{1+t} + \frac{E((b-b_t)a_t)}{t(t+1)}. \end{aligned}$$

Letting $s_t = \left|\frac{E(a_t)}{t} - \frac{c}{1+b}\right|$, the triangle inequality now gives:

$$s_{t+1} \leq s_t \left|1 - \frac{1+b}{t+1}\right| + \left|\frac{c_t - c}{t+1}\right| + \left|\frac{E[(b-b_t)a_t]}{t(t+1)}\right|.$$

Using the fact that $\lim_{t\to\infty} E(b_t) = b$, $\lim_{t\to\infty} c_t = c$, $|a_t/t| \leq C$, $\forall \epsilon > 0$, when t is large enough, we have:

$$\left|\frac{c_t - c}{t+1}\right| + \left|\frac{E[(b-b_t)a_t]}{t(t+1)}\right| < \frac{(1+b)\epsilon}{t+1}.$$

For fixed $\epsilon > 0$, when t is large enough, we have

$$s_{t+1} - \epsilon < (s_t - \epsilon)\left(1 - \frac{1+b}{t+1}\right), \left(1 - \frac{1+b}{t+1}\right) > 0.$$

Since $0 < 1 + b < \infty$, it is not difficult to see that $\prod_{t=1}^{\infty}\left(1 - \frac{1+b}{t+1}\right) \to 0$. Then for fixed $\epsilon > 0$:

$$0 \leq \lim_{t\to\infty} s_t < \epsilon.$$

ϵ can be chosen arbitrarily, so we can get $\lim_{t\to\infty} s_t \to 0$ as $\epsilon \to 0$ as desired. Therefore, we have proved that

$$\lim_{t\to\infty} \frac{E(a_t)}{t} = \frac{c}{1+b}.$$

□

Proof of Theorem 3. Discuss according to degree d.

$\forall d \geq 2$, for each time t, let $x = \frac{m_k^d(t-1)\times d}{2(t-1)+n_0}$, $y = \frac{m_k^{d-1}(t-1)\times(d-1)}{2(t-1)+n_0}$, $z = \frac{D_k(t-1)}{2(t-1)+n_0}$. Consider analyzing $E[m_k^d(t)|G(t-1)]$ in detail:

For the vertex-step at time t,

- if the new node is connected to the old node with degree $d-1$ from group k, then $m_k^d(t)$ will be $m_k^d(t-1)+1$. The probability times $m_k^d(t)$ is:

$$q\left[p_k(1-z)(1-\theta)\frac{y}{z} + p_k y + (1-p_k)y\theta\right] \times [m_k^d(t-1)+1]. \quad \text{(S5.20)}$$

- if the new node is connected to the old node with degree d from group k, then $m_k^d(t)$ will be $m_k^d(t-1)-1$. The probability times $m_k^d(t)$ is:

$$q\left[p_k(1-z)(1-\theta)\frac{x}{z}+p_k x+(1-p_k)x\theta\right]\times[m_k^d(t-1)-1]. \quad \text{(S5.21)}$$

- $m_k^d(t)$ is the same as $m_k^d(t-1)$. The probability times $m_k^d(t)$ is:

$$q\left[1-p_k(1-z)(1-\theta)\frac{x+y}{z}-p_k(x+y)-(1-p_k)(x+y)\theta\right]\times m_k^d(t-1). \quad \text{(S5.22)}$$

For the edge-step at time t,

- if the new edge's two nodes with degree $d-1$ are from group k, then $m_k^d(t)$ will be $m_k^d(t-1)+2$. The probability times $m_k^d(t)$ is:

$$(1-q)y[y+(1-z)(1-\theta)\frac{y}{z}]\times[m_k^d(t-1)+2]. \quad \text{(S5.23)}$$

- if the new edge's two nodes with degree d are from group k, then $m_k^d(t)$ will be $m_k^d(t-1)-2$. The probability times $m_k^d(t)$ is:

$$(1-q)x[x+(1-z)(1-\theta)\frac{x}{z}]\times[m_k^d(t-1)-2]. \quad \text{(S5.24)}$$

- if one vertex of the new edge is from group k with degree $d-1$, and the other is neither from group k with degree d nor degree $d-1$, then $m_k^d(t)$ will be $m_k^d(t-1)+1$. The probability times $m_k^d(t)$ is:

$$(1-q)[2y(1-z)\theta+2(z-y-x)(y+(1-z)(1-\theta)\frac{y}{z})]\times[m_k^d(t-1)+1]. \quad \text{(S5.25)}$$

- if one vertex of the new edge is from group k with degree d, and the other is neither from group k with degree d nor degree $d-1$, then $m_k^d(t)$ will be $m_k^d(t-1)-1$. The probability times $m_k^d(t)$ is:

$$(1-q)[2x(1-z)\theta+2(z-y-x)(x+(1-z)(1-\theta)\frac{x}{z})]\times[m_k^d(t-1)-1]. \tag{S5.26}$$

- if the $m_k^d(t)$ is the same as $m_k^d(t-1)$. The probability times $m_k^d(t)$ is:

$$(1-q)[1-(x^2+y^2)(1+\frac{(1-z)(1-\theta)}{z})-2(y+x)(1-z)\theta$$
$$-2(y+x)(z-y-x)(1+\frac{(1-z)(1-\theta)}{z})]\times m_k^d(t-1). \tag{S5.27}$$

By summing Equations (S5.20)–(S5.27):

$$\begin{aligned}
E[m_k^d(t)|G(t-1)] = \quad & m_d^k(t-1)+q[p_k(y-x)(1+(1-z)(1-\theta)\frac{1}{z})+(1-p_k)\theta(y-x)] \\
& +2(1-q)[(y^2-x^2)(1+(1-z)(1-\theta)\frac{1}{z})] \\
& +(1-q)[2(y-x)(1-z)\theta+2(y-x)(z-y-x)(1+(1-z)(1-\theta)\frac{1}{z})] \\
= \quad & m_k^d(t-1)+q\left[p_k(y-x)\frac{(1-\theta)}{z}+\theta(y-x)\right]+2(1-q)(y-x) \\
= \quad & m_k^d(t-1)\left[1-\frac{qdp_k(1-\theta)}{D_k(t-1)}-\frac{d(2-(2-\theta)q)}{2(t-1)+n_0}\right] \\
& +m_k^{d-1}(t-1)\left[\frac{q(d-1)p_k(1-\theta)}{D_k(t-1)}+\frac{(d-1)(2-(2-\theta)q)}{2(t-1)+n_0}\right].
\end{aligned}$$

If we take the expectation of both sides, we get the following recurrence formula.

$$E[m_k^d(t)] = E\left[m_k^d(t-1)[1 - \frac{qdp_k(1-\theta)}{D_k(t-1)} - \frac{d(2-(2-\theta)q)}{2(t-1)+n_0}]\right] + E\left[m_k^{d-1}(t-1)[\frac{q(d-1)p_k(1-\theta)}{D_k(t-1)} + \frac{(d-1)(2-(2-\theta)q)}{2(t-1)+n_0}]\right]. \tag{S5.28}$$

When $d = 1$, the formula is different. Let $x = \frac{m_k^1(t-1)}{2(t-1)+n_0}$, $z = \frac{D_k(t-1)}{2(t-1)+n_0}$.

For the vertex-step at time t,

- if the new node from group k is connected to an old node with degree > 1 or from another group, then $m_k^1(t)$ will be $m_k^1(t-1) + 1$. The probability times $m_k^1(t)$ is:

$$q\left[p_k(1-z)\theta + p_k(1-z)(1-\theta)(z-x)\frac{1}{z} + p_k(z-x)\right] \times [m_k^1(t-1)+1]. \tag{S5.29}$$

- if the new node from other groups contacts an old node with degree 1 from group k, then $m_k^1(t)$ will be $m_k^1(t-1) - 1$. The probability times $m_k^1(t)$ is:

$$q[(1-p_k)x\theta] \times [m_k^1(t-1) - 1]. \tag{S5.30}$$

- if the $m_k^1(t)$ is the same as $m_k^1(t-1)$. The probability times $m_k^1(t)$ is:

$$q\left[1 - (p_k(1-z)\theta + p_k(z-x)(1-z)(1-\theta)\frac{1}{z} + p_k(z-x)) - (1-p_k)x\theta\right] \times m_k^1(t-1). \tag{S5.31}$$

For the edge-step at time t,

- If the new edge connects both two nodes with degree 1 from group k, then $m_k^1(t)$ will be $m_k^1(t-1)-2$. The probability times $m_k^1(t)$ is:

$$(1-q)\left[x(x+(1-z)(1-\theta)\frac{x}{z})\right]\times[m_k^1(t-1)-2]. \tag{S5.32}$$

- If one node of the new edge is from group k with degree 1, the other is not, then $m_k^1(t)$ will be $m_k^1(t-1)-1$. The probability times $m_k^1(t)$ is:

$$(1-q)\left[2x(1-z)\theta+2x((z-x)+(1-z)(1-\theta)\frac{(z-x)}{z})\right]\times[m_k^1(t-1)-1]. \tag{S5.33}$$

- If the $m_k^1(t)$ is the same with $m_k^1(t-1)$, the probability times $m_k^1(t)$ is:

$$\begin{aligned}&(1-q)[(1-x^2(1+(1-(1-z)\theta)\frac{1}{z})-2x(1-z)\theta\\&-2x(z-x)(1+(1-z)(1-\theta)\frac{1}{z}))]\times m_k^1(t-1).\end{aligned} \tag{S5.34}$$

By summing Equations (S5.29)–(S5.34):

$$\begin{aligned}E[m_k^1(t)|G(t-1)] = \quad & m_k^1(t-1)+[qp_k+2(q-1)x][(1-z)\theta+(z-x)(1-z)(1-\theta)\frac{1}{z}+(z-x)]\\&-q(1-p_k)x\theta-2(1-q)[(x^2(1+(1-z)(1-\theta)\frac{1}{z})]\\ = \quad & m_k^1(t-1)+[qp_k+2(q-1)x][1-\frac{x}{z}(1-\theta)-x\theta]\\&-qx\theta+qp_kx\theta+2(q-1)x[x(\frac{1-\theta}{z}+\theta)]\\ = \quad & m_k^1(t-1)+[qp_k+2(q-1)x]-qx\theta-qp_k\frac{x}{z}(1-\theta)\\ = \quad & m_k^1(t-1)\left[1-\frac{[2-(2-\theta)q]}{2(t-1)+n_0}-\frac{qp_k(1-\theta)}{D_k(t-1)}\right]+qp_k.\end{aligned}$$

Thus, taking the expectation of both sides

$$E[m_k^1(t)] = E\left[m_k^1(t-1)\left(1 - \frac{2-(2-\theta)q}{2(t-1)+n_0} - \frac{qp_k(1-\theta)}{D_k(t-1)}\right)\right] + qp_k. \tag{S5.35}$$

By Theorem 1, we get $\left[\frac{2-(2-\theta)q}{2(t-1)+n_0} + \frac{qp_k(1-\theta)}{D_k(t-1)}\right] \times \left[(t-1)+\frac{n_0}{2}\right] \xrightarrow{a.s.} 1-\frac{q}{2}$. By Lemma 1 we have:

$$M_k^1 = \lim_{t\to\infty} \frac{E[m_k^1(t)]}{t} = \frac{2qp_k}{4-q}.$$

Next, consider $d \geq 2$, by Equation (S5.28) and Lemma 1:

$$M_k^d = \lim_{t\to\infty} \frac{E(m_k^d(t))}{t} = \frac{(d-1)(2-q)}{2+d(2-q)} M_k^{d-1}.$$

Thus, $\forall d \geq 2$, we can write,

$$M_k^d = \frac{2qp_k}{4-q} \prod_{j=2}^{d} \frac{(j-1)(2-q)}{2+j(2-q)}.$$

Then we prove the power-law, $\forall d \geq 2$

$$\frac{M_k^d}{M_k^{d-1}} = \frac{(d-1)(2-q)}{2+d(2-q)} = 1 - \frac{2+(2-q)}{2+d(2-q)} = 1 - \frac{2/(2-q)+1}{d} + O(\frac{1}{d^2}).$$

Consider if $M_k^d \propto d^{-\beta_k}$. Then

$$\frac{M_k^d}{M_k^{d-1}} = \frac{d^{-\beta_k}}{(d-1)^{-\beta_k}} = (1-\frac{1}{d})^{\beta_k} = 1 - \frac{\beta_k}{d} + O(\frac{1}{d^2}).$$

Thus, the β_k for M_k^d is $1 + 2/(2-q)$. □

Proof of Theorem 4. Consider the likelihood function of $\{G(t)\}_{t=0}^{T}$, let $P_k(t) = \frac{D_k(t-1)}{2(t-1)+n_0}$:

$$
\begin{aligned}
& L(\psi|\{G(t)\}_{t=0}^{T}) \\
= \quad & \prod_{k=1}^{K}\prod_{t=1}^{T}\left[qp_k\left(P_k(t)+(1-\theta)\left(1-P_k(t)\right)\right)\right]^{\mathbb{1}\{v(t)=1,g_{e_1(t)}=g_{e_2(t)}=k\}} \\
& \times\prod_{k=1}^{K}\prod_{t=1}^{T}\left[qp_k\left(1-P_k(t)\right)\theta\right]^{\mathbb{1}\{v(t)=1,g_{e_1(t)}=k,g_{e_2(t)}\neq k\}} \\
& \times\prod_{k=1}^{K}\prod_{t=1}^{T}\left[(1-q)P_k(t)\left(P_k(t)+(1-\theta)\left(1-P_k(t)\right)\right)\right]^{\mathbb{1}\{v(t)=0,g_{e_1(t)}=g_{e_2(t)}=k\}} \\
& \times\prod_{k=1}^{K}\prod_{t=1}^{T}\left[(1-q)P_k(t)\left(1-P_k(t)\right)\theta\right]^{\mathbb{1}\{v(t)=0,g_{e_1(t)}=k,g_{e_2(t)}\neq k\}}.
\end{aligned}
$$

The log likelihood function is

$$
\begin{aligned}
& \log L(\psi|\{G(t)\}_{t=0}^{T}) \\
= \quad & \sum_{t=1}^{T}\left[\mathbb{1}\{v(t)=1\}\log(q)+\mathbb{1}\{v(t)=0\}\log(1-q)\right]+\sum_{k=1}^{K}\sum_{t=1}^{T}\mathbb{1}\{v(t)=1,g_{e_1(t)}=k\}\log(p_k) \\
& +\sum_{k=1}^{K}\sum_{t=1}^{T}\mathbb{1}\{v(t)=1,g_{e_1(t)}=g_{e_2(t)}=k\}\log\left(P_k(t)+(1-\theta)\left(1-P_k(t)\right)\right) \\
& +\sum_{k=1}^{K}\sum_{t=1}^{T}\mathbb{1}\{v(t)=1,g_{e_1(t)}=k,g_{e_2(t)}\neq k\}\log\left(\left(1-P_k(t)\right)\theta\right) \\
& +\sum_{k=1}^{K}\sum_{t=1}^{T}\mathbb{1}\{v(t)=0,g_{e_1(t)}=g_{e_2(t)}=k\}\log\left(P_k(t)\left(P_k(t)+(1-\theta)\left(1-P_k(t)\right)\right)\right) \\
& +\sum_{k=1}^{K}\sum_{t=1}^{T}\mathbb{1}\{v(t)=0,g_{e_1(t)}=k,g_{e_2(t)}\neq k\}\log\left(P_k(t)\left(1-P_k(t)\right)\theta\right). \qquad \text{(S5.36)}
\end{aligned}
$$

Since $\sum_{k=1}^{K}p_k = 1$, we can rewrite $p_K = 1-\sum_{k=1}^{K-1}p_k$. The score functions of Equation (S5.36) for p_k, $k = 1, \cdots, K-1$, q are as follows:

$$\frac{\partial}{\partial p_k} \log L(\psi | \{G(t)\}_{t=0}^{T}) = \frac{\sum_{t=1}^{T} \mathbb{1}\{v(t)=1, g_{e_1(t)}=k\}}{p_k} - \frac{\sum_{t=1}^{T} \mathbb{1}\{v(t)=1, g_{e_1(t)}=K\}}{1-\sum_{k=1}^{K-1} p_k}. \tag{S5.37}$$

$$\frac{\partial}{\partial q} \log L(\psi | \{G(t)\}_{t=0}^{T}) = \frac{\sum_{t=1}^{T} \mathbb{1}\{v(t)=1\}}{q} - \frac{\sum_{t=1}^{T} \mathbb{1}\{v(t)=0\}}{1-q}. \tag{S5.38}$$

Equations (S5.37) and (S5.38) are monotonically decreasing continuous functions of $p_k \in [0,1]$ and $q \in [0,1]$.

What is more, when $p_k = 0, q = 0$, Equations (S5.37) and (S5.38) both go to $+\infty$; and when $p_k = 1, q = 1$, Equations (S5.37) and (S5.38) both go to $-\infty$. Therefore, we know that Equations (S5.37) and (S5.38) both have a unique root:

$$\hat{p}_k = \frac{\sum_{t=1}^{T} \mathbb{1}\{v(t)=1, g_{e_1(t)}=k\}}{\sum_{t=1}^{T} \mathbb{1}\{v(t)=1\}}.$$

$$\hat{q} = \frac{\sum_{t=1}^{T} \mathbb{1}\{v(t)=1\}}{T}.$$

By the law of large numbers, get

$$\begin{aligned} \hat{p}_k = & \quad \frac{\sum_{t=1}^{T} \mathbb{1}\{v(t)=1, g_{e_1(t)}=k\}}{\sum_{t=1}^{T} \mathbb{1}\{v(t)=1\}} \\ = & \quad \frac{\sum_{t=1}^{T} \mathbb{1}\{v(t)=1, g_{e_1(t)}=k\}}{q^* T} \times \frac{q^* T}{\sum_{t=1}^{T} \mathbb{1}\{v(t)=1\}} \\ \xrightarrow{a.s.} & \quad p_k^*. \end{aligned}$$

$$\hat{q} = \frac{\sum_{t=1}^{T} \mathbb{1}\{v(t)=1\}}{T} \xrightarrow{a.s.} q^*.$$

Next, considering the score function for θ, we have

$$
\begin{aligned}
& \frac{\partial}{\partial\theta}\log L(\psi|\{G(t)\}_{t=0}^{T}) \\
= \quad & \frac{\partial}{\partial\theta}\log L_2(\theta|\{G(t)\}_{t=0}^{T}) \\
= \quad & -\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{v(t)=1, g_{e_1(t)}=g_{e_2(t)}=k\}(1-P_k(t))}{P_k(t)+(1-\theta)\,(1-P_k(t))} \\
& -\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{v(t)=0, g_{e_1(t)}=g_{e_2(t)}=k\}(1-P_k(t))}{P_k(t)+(1-\theta)\,(1-P_k(t))} \\
& +\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{v(t)=1, g_{e_1(t)}=k, g_{e_2(t)}\neq k\}}{\theta} \\
& +\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{v(t)=0, g_{e_1(t)}=k, g_{e_2(t)}\neq k\}}{\theta}.
\end{aligned} \tag{S5.39}
$$

Equation (S5.39) is a strictly decreasing continuous function of $\theta \in (0, 1+\epsilon)$, where $\epsilon = \min\{\min\{P_k(t)/(1-P_k(t)) : t \in [1,T], g_{e_1(t)} = g_{e_2(t)} = k\} : k \in [K]\}$. Equation (S5.39) $\to \infty$ when $\theta \to 0$ and Equation (S5.39) $\to -\infty$ when $\theta \to 1+\epsilon$. Thus, there is a unique root for $\frac{\partial}{\partial\theta}\log L(\psi|\{G(t)\}_{t=0}^{T})$ (maximum point for $\log L_2(\theta|\{G(t)\}_{t=0}^{T})$) of $\theta \in (0, 1+\epsilon)$.

Let $\hat{\theta} = \arg\max_{\theta\in(0,1+\epsilon)}\log L_2(\theta|\{G(t)\}_{t=0}^{T})$. $\hat{\theta}$ is the unique maximum point for $\log L_2(\theta|\{G(t)\}_{t=0}^{T})$. Thus, $\forall\theta \neq \hat{\theta} \in (0, 1+\epsilon)$, $\log L_2(\hat{\theta}|\{G(t)\}_{t=0}^{T}) > \log L_2(\theta|\{G(t)\}_{t=0}^{T})$. And it implies

$$
\frac{1}{T}\log L_2(\hat{\theta}|\{G(t)\}_{t=0}^{T}) > \frac{1}{T}\log L_2(\theta|\{G(t)\}_{t=0}^{T}). \tag{S5.40}
$$

We have

$$
\begin{aligned}
& \frac{\log L_2(\theta|\{G(t)\}_{t=0}^{T})}{T} \\
= \quad & \sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{v(t)=1, g_{e_1(t)}=g_{e_2(t)}=k\}\log\left(P_k(t)+(1-\theta)\left(1-P_k(t)\right)\right)}{T} \\
& +\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{v(t)=1, g_{e_1(t)}=k, g_{e_2(t)}\neq k\}\log\left(\left(1-P_k(t)\right)\theta\right)}{T} \\
& +\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{v(t)=0, g_{e_1(t)}=g_{e_2(t)}=k\}\log\left(P_k(t)\left(P_k(t)+(1-\theta)\left(1-P_k(t)\right)\right)\right)}{T} \\
& +\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{v(t)=0, g_{e_1(t)}=k, g_{e_2(t)}\neq k\}\log\left(P_k(t)\left(1-P_k(t)\right)\theta\right)}{T} \\
= \quad & \sum_{k=1}^{K}\frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=1, g_{e_1(t)}=g_{e_2(t)}=k\}\log\left(P_k(t)+(1-\theta)\left(1-P_k(t)\right)\right)}{\sum_{t=1}^{T}\mathbb{1}\{v(t)=1, g_{e_1(t)}=g_{e_2(t)}=k\}} \\
& \times\frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=1, g_{e_1(t)}=g_{e_2(t)}=k\}}{T} \\
& +\sum_{k=1}^{K}\frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=1, g_{e_1(t)}=k, g_{e_2(t)}\neq k\}\log\left(\left(1-P_k(t)\right)\theta\right)}{\sum_{t=1}^{T}\mathbb{1}\{v(t)=1, g_{e_1(t)}=k, g_{e_2(t)}\neq k\}} \\
& \times\frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=1, g_{e_1(t)}=k, g_{e_2(t)}\neq k\}}{T} \\
& +\sum_{k=1}^{K}\frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=0, g_{e_1(t)}=g_{e_2(t)}=k\}\log\left(P_k(t)\left(P_k(t)+(1-\theta)\left(1-P_k(t)\right)\right)\right)}{\sum_{t=1}^{T}\mathbb{1}\{v(t)=0, g_{e_1(t)}=g_{e_2(t)}=k\}} \\
& \times\frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=0, g_{e_1(t)}=g_{e_2(t)}=k\}}{T} \\
& +\sum_{k=1}^{K}\frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=0, g_{e_1(t)}=k, g_{e_2(t)}\neq k\}\log\left(P_k(t)\left(1-P_k(t)\right)\theta\right)}{\sum_{t=1}^{T}\mathbb{1}\{v(t)=0, g_{e_1(t)}=k, g_{e_2(t)}\neq k\}} \\
& \times\frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=0, g_{e_1(t)}=k, g_{e_2(t)}\neq k\}}{T}. \qquad \text{(S5.41)}
\end{aligned}
$$

By LLN and Theorem 1 in the main paper, get

$$\begin{aligned} & \lim_{T\to\infty} \frac{\log L_2(\theta|\{G_t\}_{t=0}^T)}{T} \\ = \quad & \sum_{k=1}^{K} \log(p_k^* + (1-\theta)(1-p_k^*)) \times [(p_k^*)^2 + p_k^*(1-p_k^*)(1-\theta^*)] \\ & + \sum_{k=1}^{K} \log\left((1-p_k^*)\,\theta\right) \times p_k^*(1-p_k^*)\theta^*. \end{aligned} \tag{S5.42}$$

Consider a random variable $Y = (x, z)$ such that $x \in \{1, \cdots, K\}$ and $z \in \{0, 1\}$. The probability distribution of Y is

$$P(y;\theta) = \begin{cases} p_k^*(1-p_k^*)\theta, & \text{if } x = k, y = 0; \\ (p_k^*)^2 + p_k^*(1-p_k^*)(1-\theta), & \text{if } x = k, y = 1. \end{cases} \tag{S5.43}$$

For any other $\theta \neq \theta^* \in (0, 1+\epsilon^*)$, where $\epsilon^* = \min_{k\in[K]}\{p_k^*/(1-p_k^*)\}_k$.

$$\begin{aligned} E_{\theta^*} \log[P(Y;\theta^*)/P(Y;\theta)] = \quad & -E_{\theta^*} \log[P(Y;\theta)/P(Y;\theta^*)] \\ \geq \quad & -\log[E_{\theta^*}(P(Y;\theta)/f(Y;\theta^*))] \\ = \quad & -\log[1] = 0. \end{aligned} \tag{S5.44}$$

Equation (S5.44) implies that

$$\forall\theta \neq \theta^* \in (0, 1+\epsilon^*),\ E_{\theta^*} \log[P(Y;\theta^*)] > E_{\theta^*} \log[P(Y;\theta)]. \tag{S5.45}$$

Moreover, we can prove $E_{\theta^*} \log[P(Y;\theta^*)] = \lim_{T\to\infty} \frac{\log L_2(\theta^*|\{G_t\}_{t=0}^T)}{T} + \sum_{k=1}^K p_k^* \log(p_k^*)$. Thus, by Equation (S5.45), when T is large enough, we know $\frac{\log L_2(\theta^*|\{G_t\}_{t=0}^T)}{T} > \frac{\log L_2(\theta|\{G_t\}_{t=0}^T)}{T}$, $\forall\theta \neq \theta^* \in (0, 1+\epsilon^*)$. Combining this result with Equation (S5.40), get $P(\lim_{T\to\infty}\hat{\theta}=\theta^*)=1$, which means

$$\lim_{T\to\infty}\hat{\theta}\xrightarrow{a.s.}\theta^*.$$

□

Proof of Theorem 5. Let $l(\theta|\{G(t)\}_{t=0}^T) = \frac{\partial \log L_2(\theta|\{G(t)\}_{t=0}^T)}{\partial\theta}$ and $P_k(t) = \frac{D_k(t-1)}{2(t-1)+n_0}$. By Equation (S5.39), we get

$$\begin{aligned}
l(\theta|\{G(t)\}_{t=0}^T) = \quad & -\sum_{k=1}^K\sum_{t=1}^T \frac{\mathbb{1}\{v(t)=1, g_{e_1(t)}=g_{e_2(t)}=k\}(1-P_k(t))}{P_k(t)+(1-\theta)\,(1-P_k(t))} \\
& -\sum_{k=1}^K\sum_{t=1}^T \frac{\mathbb{1}\{v(t)=0, g_{e_1(t)}=g_{e_2(t)}=k\}(1-P_k(t))}{P_k(t)+(1-\theta)\,(1-P_k(t))} \\
& +\sum_{k=1}^K\sum_{t=1}^T \frac{\mathbb{1}\{v(t)=1, g_{e_1(t)}=k, g_{e_2(t)}\neq k\}}{\theta} \\
& +\sum_{k=1}^K\sum_{t=1}^T \frac{\mathbb{1}\{v(t)=0, g_{e_1(t)}=k, g_{e_2(t)}\neq k\}}{\theta}.
\end{aligned}$$

Further, we can get:

$$\begin{aligned}
l'(\theta|\{G(t)\}_{t=0}^T) = \quad & \frac{\partial l'(\theta|\{G(t)\}_{t=0}^T)}{\partial\theta} \\
= \quad & -\sum_{k=1}^K\sum_{t=1}^T \frac{\mathbb{1}\{v(t)=1, g_{e_1(t)}=g_{e_2(t)}=k\}\,(1-P_k(t))^2}{[P_k(t)+(1-\theta)\,(1-P_k(t))]^2} \\
& -\sum_{k=1}^K\sum_{t=1}^T \frac{\mathbb{1}\{v(t)=0, g_{e_1(t)}=g_{e_2(t)}=k\}\,(1-P_k(t))^2}{[P_k(t)+(1-\theta)\,(1-P_k(t))]^2} \\
& -\sum_{k=1}^K\sum_{t=1}^T \frac{\mathbb{1}\{v(t)=1, g_{e_1(t)}=k, g_{e_2(t)}\neq k\}}{\theta^2} \\
& -\sum_{k=1}^K\sum_{t=1}^T \frac{\mathbb{1}\{v(t)=0, g_{e_1(t)}=k, g_{e_2(t)}\neq k\}}{\theta^2}.
\end{aligned}$$

Let θ^* be the true parameter for the network $\{G(t)\}_{t=0}^T$ by a KPA model, and taking the Taylor Expansion of $l(\theta|\{G(t)\}_{t=0}^T)$ around θ^*, we can get:

$$l(\theta|\{G(t)\}_{t=0}^T) = l(\theta^*|\{G(t)\}_{t=0}^T) + (\theta-\theta^*)l'(\theta^*|\{G(t)\}_{t=0}^T) + (\theta-\theta^*)^2 l'(\theta^{**}|\{G(t)\}_{t=0}^T), \tag{S5.46}$$

where $\theta^{**} = \theta^* + \xi \circ (\theta - \theta^*)$, $\xi \in [0,1]$. As a result of $l(\hat{\theta}|\{G(t)\}_{t=0}^T) = 0$, Equation (S5.46) implies

$$T^{1/2}(\hat{\theta}-\theta^*) = -\frac{\frac{T^{1/2} l(\theta^*|\{G(t)\}_{t=0}^T)}{T}}{\frac{l'(\theta^*|\{G(t)\}_{t=0}^T)}{T} + \frac{l'(\theta^{**}|\{G(t)\}_{t=0}^T)(\hat{\theta}-\theta^*)}{T}}. \tag{S5.47}$$

Consider the random variable Y and its probability distribution $P(\cdot;\theta^*)$ defined by Equation (S5.43).

$$\begin{aligned} I(\theta^*) = \quad & E_{\theta^*}\left[-\frac{\partial^2 \log P(Y;\theta^*)}{\partial(\theta^*)^2}\right] = E_{\theta^*}\left[\frac{P'(Y;\theta^*)^2}{P(Y;\theta^*)^2}\right] \\ = \quad & \sum_y \frac{P'(y;\theta^*)^2}{P(y;\theta^*)} = \sum_{k=1}^K \left[\frac{p_k^*(1-p_k^*)}{\theta^*} + \frac{p_k^*(1-p_k^*)^2}{p_k^* + (1-p_k^*)(1-\theta^*)}\right]. \end{aligned} \tag{S5.48}$$

By the martingale convergence theorem, martingale difference central limit theorem and Equation (S5.48), we have that when T tends to infinity,

$$\hat{\theta} \xrightarrow{a.s.} \theta^*, \quad \frac{l(\theta^*|\{G(t)\}_{t=0}^T)}{T^{1/2}} \xrightarrow{d} N(0, I(\theta^*)), \tag{S5.49}$$

$$-\frac{l'(\theta^*|\{G(t)\}_{t=0}^T)}{T} \xrightarrow{a.s.} I(\theta^*), \tag{S5.50}$$

and

$$-\frac{l'(\theta^{**}|\{G(t)\}_{t=0}^{T})}{T} \xrightarrow{a.s.} \sum_{k=1}^{K}\left[\frac{\theta^* p_k^*(1-p_k^*)}{(\theta^{**})^2} + \frac{(p_k^*+(1-p_k^*)(1-\theta^*))p_k^*(1-p_k^*)^2}{(p_k^*+(1-p_k^*)(1-\theta^{**}))^2}\right]. \tag{S5.51}$$

By Equations (S5.49)–(S5.51), it follows that when T tends to infinity,

$$T^{1/2}(\hat{\theta}-\theta^*) \xrightarrow{d} N(0, \frac{1}{I(\theta^*)}).$$

Let $\mathbf{p} = (p_1, \cdots, p_{K-1})'$ and $p_K = 1 - \sum_{k=1}^{K-1} p_k$, we define

$$\log L_1(\mathbf{p}|\{G(t)\}_{t=0}^{T}) = \sum_{k=1}^{K}\sum_{t=1}^{T} \mathbb{1}\{v(t)=1, g_{e_1(t)}=k\}\log(p_k).$$

We also define

$$l(\mathbf{p}|\{G(t)\}_{t=0}^{T}) = \frac{\partial \log L_1(\mathbf{p}|\{G(t)\}_{t=0}^{T})}{\partial \mathbf{p}} = \begin{pmatrix} \frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=1,g_{e_1(t)}=1\}}{p_1} - \frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=1,g_{e_1(t)}=K\}}{1-\sum_{k=1}^{K-1}p_k} \\ \vdots \\ \frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=1,g_{e_2(t)}=K-1\}}{p_{K-1}} - \frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=1,g_{e_1(t)}=K\}}{1-\sum_{k=1}^{K-1}p_k} \end{pmatrix}.$$

Further, let $p_K = 1 - \sum_{k=1}^{K-1} p_k$:

$$l'(\mathbf{p}|\{G(t)\}_{t=0}^{T}) = \frac{\partial \log l(\mathbf{p}|\{G(t)\}_{t=0}^{T})}{\partial \mathbf{p}}$$

$$= \begin{bmatrix} -\sum_{t=1}^{T}[\frac{\mathbb{1}\{v(t)=1,g_{e_1(t)}=1\}}{p_1^2} + \frac{\mathbb{1}\{v(t)=1,g_{e_1(t)}=K\}}{p_K^2}] & \cdots & -\frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=1,g_{e_1(t)}=K\}}{p_K^2} \\ \vdots & & \vdots \\ -\frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=1,g_{e_1(t)}=K\}}{p_K^2} & \cdots & -\sum_{t=1}^{T}[\frac{\mathbb{1}\{v(t)=1,g_{e_1(t)}=K-1\}}{p_{K-1}^2} + \frac{\mathbb{1}\{v(t)=1,g_{e_1(t)}=K\}}{p_K^2}] \end{bmatrix}.$$

Let $\mathbf{p}^*$ be the true parameters for the KPA model of $\{G(t)\}_{t=0}^{T}$. Then, using a multivariate Taylor expansion of $l(\mathbf{p}|\{G(t)\}_{t=0}^{T})$ round $\mathbf{p}^*$, we can get:

$$l(\mathbf{p}|\{G(t)\}_{t=0}^{T}) = l(\mathbf{p}^*|\{G(t)\}_{t=0}^{T}) + l'(\mathbf{p}^{**}|\{G(t)\}_{t=0}^{T})(\mathbf{p}-\mathbf{p}^*), \quad \text{(S5.52)}$$

where $\mathbf{p}^{**} = \mathbf{p}^* + \boldsymbol{\xi} \circ (\mathbf{p} - \mathbf{p}^*)$, $\boldsymbol{\xi} \in [0,1]^{K-1}$.

Set $\mathbf{p} = \hat{\mathbf{p}}$ by the MLE, $\mathbf{p}^{**} = \mathbf{p}^* + \boldsymbol{\xi} \circ (\hat{\mathbf{p}} - \mathbf{p}^*)$. According to Theorem 4 in the main paper and by martingale difference central limit theorem, when T goes to infinity,

$$\mathbf{p}^{**} \xrightarrow{a.s.} \mathbf{p}^*, \qquad \frac{l(\mathbf{p}^*|\{G(t)\}_{t=0}^{T})}{T^{1/2}} \xrightarrow{d} N(0, I(\mathbf{p}^*)), \quad \text{(S5.53)}$$

and

$$-\frac{l'(\mathbf{p}^{**}|\{G(t)\}_{t=0}^{T})}{T} \xrightarrow{a.s.} I(\mathbf{p}^*), \quad \text{(S5.54)}$$

where set $p_K^* = 1 - \sum_{k=1}^{K-1} p_k^*$:

$$I(\mathbf{p}^*) = q^* \times \begin{bmatrix} \frac{p_1^*+p_K^*}{p_1^* p_K^*} & \frac{1}{p_K^*} & \cdots & \frac{1}{p_K^*} \\ \frac{1}{p_K^*} & \frac{p_2^*+p_K^*}{p_2^* p_K^*} & \cdots & \frac{1}{p_K^*} \\ \vdots & & \cdots & \vdots \\ \frac{1}{p_K^*} & \cdots & \frac{1}{p_K^*} & \frac{p_{K-1}^*+p_K^*}{p_{K-1}^* p_K^*} \end{bmatrix}. \quad \text{(S5.55)}$$

By Equations (S5.52)–(S5.55), then when T goes to infinity,

$$\sqrt{T}(\hat{\mathbf{p}} - \mathbf{p}^*)' \xrightarrow{d} N(0, \frac{1}{I(\mathbf{p}^*)}).$$

$\forall q \in (0,1]$, we have

$$\log L_3(q|\{G(t)\}_{t=0}^{T}) = \sum_{t=1}^{T} [\mathbb{1}\{v(t)=1\} \log(q) + \mathbb{1}\{v(t)=0\} \log(1-q)].$$

$$l(q|\{G(t)\}_{t=0}^{T}) = \frac{\partial}{\partial q}\log L(\psi|\{G(t)\}_{t=0}^{T}) = \frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=1\}}{q} - \frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=0\}}{1-q}.$$

$$l'(q|\{G(t)\}_{t=0}^{T}) = \frac{\partial}{\partial q}l'(q|\{G(t)\}_{t=0}^{T}) = -\frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=1\}}{q^2} - \frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=0\}}{(1-q)^2}.$$

Let q^* be the true parameter, and using a Taylor expansion of $l(q|\{G(t)\}_{t=0}^{T})$ around q^*, we can get:

$$l(q|\{G(t)\}_{t=0}^{T}) = l(q^*|\{G(t)\}_{t=0}^{T}) + (q-q^*)l'(q^*|\{G(t)\}_{t=0}^{T}) + (q-q^*)^2 l'(q^{**}|\{G(t)\}_{t=0}^{T}). \tag{S5.56}$$

By Theorem 4 in the main paper and martingale difference central limit theorem, when T goes to infinity,

$$\frac{l(q^*|\{G(t)\}_{t=0}^{T})}{T^{1/2}} \xrightarrow{d} N(0, \frac{1}{q^*(1-q^*)}), \quad \hat{q} \xrightarrow{a.s.} q^*, \tag{S5.57}$$

$$\begin{aligned}\frac{l'(q^{**}|\{G(t)\}_{t=0}^{T})}{T} &= -\frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=1\}}{(q^{**})^2 T} - \frac{\sum_{t=1}^{T}\mathbb{1}\{v(t)=0\}}{(1-q^{**})^2 T} \\ &\xrightarrow{a.s.} -\frac{q^*}{(q^{**})^2} - \frac{1-q^*}{(1-q^{**})^2},\end{aligned} \tag{S5.58}$$

and

$$-\frac{l'(q^*|\{G_t\}_{t=0}^{T})}{T} \xrightarrow{a.s.} \frac{1}{q^*(1-q^*)}. \tag{S5.59}$$

Let $q = \hat{q}$, $l(\hat{q}|\{G(t)\}_{t=0}^{T}) = 0$, we can get

$$(\hat{q} - q^*) = -\frac{l(q^*|\{G(t)\}_{t=0}^{T})}{l'(q^*|\{G_t\}_{t=0}^{T}) + (q-q^*)l'(q^{**}|\{G(t)\}_{t=0}^{T})}. \tag{S5.60}$$

by Equations (S5.56)–(S5.60), we have that when T goes to infinity,

$$T^{1/2}(\hat{q} - q^*) \xrightarrow{d} N(0, q^*(1-q^*)).$$

□

Proof of Theorem 6. By Slutsky's theorem, $\forall k \in \{1, \cdots, K\}$,

$$\begin{aligned}
\tilde{p}_k = & \quad \frac{|\mathcal{V}_k(T)|}{|\mathcal{V}(T)|} = \frac{n_{0,k} + \sum_{t=1}^T \mathbb{1}\{v(t) = 1, g_{e_1(t)} = k\}}{n_0 + \sum_{t=1}^T \mathbb{1}\{v(t) = 1\}} \\
= & \quad \frac{n_{0,k}}{n_0 + \sum_{t=1}^T \mathbb{1}\{v(t) = 1\}} + \hat{p}_k \\
\xrightarrow{a.s.} & \quad p_k^*. \\
\tilde{q} = & \quad \frac{|\mathcal{V}(T)|}{|\mathcal{E}(T)|} = \frac{n_0 + \sum_{t=1}^T \mathbb{1}\{v(t) = 1\}}{e_0 + T} \\
= & \quad \frac{T}{e_0 + T}\hat{q} + \frac{n_0}{e_0 + T} \\
\xrightarrow{a.s.} & \quad q^*.
\end{aligned}$$

where e_0, n_0 means the number of edges and vertices in $G(0)$, and n_0^k means the number of vertices from group k in $G(0)$.

Considering $\tilde{\theta}$, we have

$$\begin{aligned}
\frac{L_T(\theta|G_T)}{T} = & \quad \sum_{k=1}^K \left[\frac{|\mathcal{E}_{k,1}(T)|}{T} \log(\frac{D_k(T)}{2|\mathcal{E}(T)|}(\frac{D_k(T)}{2|\mathcal{E}(T)|} + (1-\theta)(1 - \frac{D_k(T)}{2|\mathcal{E}(T)|}))) \right] \\
& + \sum_{k=1}^K \left[\frac{|\mathcal{E}_{k,0}(T)|}{T} \log(\frac{D_k(T)}{2|\mathcal{E}(T)|}\theta(1 - \frac{D_k(T)}{2|\mathcal{E}(T)|})) \right].
\end{aligned}$$

By LLN and Theorem 4 in the main paper, we have

$$\begin{aligned}
& \lim_{T\to\infty} \frac{L_T(\theta|G(T))}{T} \\
\xrightarrow{a.s.} \quad & \sum_{k=1}^{K} [((p_k^*)^2 + p_k^*(1-p_k^*)(1-\theta^*)) \log((p_k^*)^2 + p_k^*(1-p_k^*)(1-\theta)) \\
& +(p_k^*(1-p_k^*)\theta^*) \log(p_k^*(1-p_k^*)\theta)] \\
= \quad & E_{\theta^*}[\log P(Y;\theta)]. \qquad \text{(S5.61)}
\end{aligned}$$

Equation (S5.61) implies $\frac{1}{T}L_T(\theta^*|G(T)) > \frac{1}{T}L_T(\theta|G(T))$, $\forall\theta \neq \theta^* \in (0, 1+\tilde{\epsilon})$ when T is large enough. So we can get $\tilde{\theta} = \arg\max_{\theta\in(0,1+\tilde{\epsilon})} L_T(\theta|G(T)) \xrightarrow{a.s.} \theta^*$, where $\tilde{\epsilon} = \min_{k\in[K]}\{\frac{D_k(T)}{2|\mathcal{E}(T)|}/(1-\frac{D_k(T)}{2|\mathcal{E}(T)|})\}_k$. □

Proof of Theorem 7.

$$\forall\varepsilon > 0,\ P\left(\frac{|\hat{\tau}-\tau^*|}{T} > \varepsilon\right) = P\left(|\hat{\tau}-\tau^*| > \varepsilon T\right).$$

Let

$\hat{\theta}_{1,\hat{\tau}} = \arg\max_{\theta\in(0,1+\hat{\epsilon}_1)} \log L_2(\theta|\{G(t)\}_{t=0}^{\hat{\tau}})$, $\hat{\theta}_{2,\hat{\tau}} = \arg\max_{\theta\in(0,1+\hat{\epsilon}_2)} \log L_2(\theta|\{G(t)\}_{t=\hat{\tau}}^{T})$, where $\hat{\epsilon}_1 = \min\{\min\{P_k(t)/(1-P_k(t)) : t \in [1,\hat{\tau}], g_{e_1(t)} = g_{e_2(t)} = k\} : k \in [K]\}$, $\hat{\epsilon}_2 = \min\{\min\{P_k(t)/(1-P_k(t)) : t \in [\hat{\tau}+1, T], g_{e_1(t)} = g_{e_2(t)} = k\} : k \in [K]\}$.

$\hat{\theta}_{1,\tau^*} = \arg\max_{\theta\in(0,1+\epsilon_1^*)} \log L_2(\theta|\{G(t)\}_{t=0}^{\tau^*})$, $\hat{\theta}_{2,\tau^*} = \arg\max_{\theta\in(0,1+\epsilon_2^*)} \log L_2(\theta|\{G(t)\}_{t=\tau^*}^{T})$, where $\epsilon_1^* = \min\{\min\{P_k(t)/(1-P_k(t)) : t \in [1,\tau^*], g_{e_1(t)} = g_{e_2(t)} = k\} : k \in [K]\}$, $\epsilon_2^* = \min\{\min\{P_k(t)/(1-P_k(t)) : t \in [\tau^*+1, T], g_{e_1(t)} = g_{e_2(t)} = k\} : k \in [K]\}$.

By Theorem 4 in the main paper, we can get $\hat{\theta}_{1,\tau^*} \xrightarrow{a.s.} \theta_1^*$, $\hat{\theta}_{2,\tau^*} \xrightarrow{a.s.} \theta_2^*$.

First, we consider the case $\hat{\tau} - \tau^* > \varepsilon T$. Divide the time range into three intervals $[0, T] = [0, \tau^*] \cup [\tau^* + 1, \hat{\tau}] \cup [\hat{\tau} + 1, T]$.

$$\frac{\log L_2(\theta|\{G(t)\}_{t=0}^{\hat{\tau}})}{\hat{\tau}} = \frac{\tau^*}{\hat{\tau}} \frac{\log L_2(\theta|\{G(t)\}_{t=0}^{\tau^*})}{\tau^*} + \frac{\hat{\tau} - \tau^*}{\hat{\tau}} \frac{\log L_2(\theta|\{G(t)\}_{t=\tau^*}^{\hat{\tau}})}{\hat{\tau} - \tau^*}.$$

Furthermore, let $\alpha = \tau^*/\hat{\tau}$, under Assumption 5 in the main paper, we have $\alpha \in [c/(1-c), 1 - \varepsilon/(1-c)]$.

Let $\epsilon^* = \min_{k \in [K]}\{p_k/(1-p_k)\}_k$. For $\forall \theta^* \in (0, 1+\epsilon^*)$,

$$\frac{\partial E_{\theta^*}[\log(P(Y;\theta))]}{\partial \theta} = -\sum_{k=1}^{K} \left[\frac{[p_k + (1-\theta^*)(1-p_k)](1-p_k)p_k}{p_k + (1-\theta)(1-p_k)} - \frac{\theta^*(1-p_k)p_k}{\theta} \right]$$

$$= \begin{cases} 0, & \text{if } \theta = \theta^*; \\ \infty, & \text{if } \theta \to 0; \\ -\infty, & \text{if } \theta \to 1 + \epsilon^*. \end{cases} \tag{S5.62}$$

For $\forall \alpha \in [c/(1-c), 1 - \varepsilon/(1-c)]$, Equation (S5.62) implies that the derivative function of $\alpha E_{\theta_1^*}[\log(P(Y;\theta))] + (1-\alpha)E_{\theta_2^*}[\log(P(Y;\theta))]$ is strictly decreasing and has a unique root. Thus, $\alpha E_{\theta_1^*}[\log(P(Y;\theta))] + (1-\alpha)E_{\theta_2^*}[\log(P(Y;\theta))]$ has a unique maximum point. Let $\theta_\alpha^* = \arg\max_{\theta \in (0, 1+\epsilon^*)}[\alpha E_{\theta_1^*}[\log(P(Y;\theta))] + (1-\alpha)E_{\theta_2^*}[\log(P(Y;\theta))]]$. And we set:

$$\Theta_0 = \left\{ \theta_\alpha^* : \alpha \in [\frac{c}{1-c}, 1 - \frac{\varepsilon}{1-c}] \right\}.$$

$\alpha \frac{\log L_2(\theta|\{G(t)\}_{t=0}^{\tau^*})}{\tau^*} + (1-\alpha)\frac{\log L_2(\theta|\{G(t)\}_{t=\tau^*}^{\hat{\tau}})}{\hat{\tau}-\tau^*}$ also has a unique maximum point by Equation (S5.39).

For $\forall\alpha \in [c/(1-c), 1-\varepsilon/(1-c)]$, let

$\hat{\theta}_\alpha = \arg\max_{\theta\in(0,1+\epsilon)} \left[\alpha\frac{\log L_2(\theta|\{G(t)\}_{t=0}^{\tau^*})}{\tau^*} + (1-\alpha)\frac{\log L_2(\theta|\{G(t)\}_{t=\tau^*}^{\hat{\tau}})}{\hat{\tau}-\tau^*}\right]$, $\epsilon = \min\{\min\{P_k(t)/(1-P_k(t)) : t \in [1,T], g_{e_1(t)} = g_{e_2(t)} = k\} : k \in [K]\}$.

By Equations (S5.42)–(S5.45), we can get $\hat{\theta}_\alpha \xrightarrow{a.s.} \theta^*_\alpha$.

Furthermore, we can get the larger α is, the closer the θ^*_α is to θ^*_1. So the distance from point θ^*_1 to set Θ_0 is $\left|\theta^*_1 - \theta^*_{\{1-\frac{\varepsilon}{1-c}\}}\right|$, the distance from point θ^*_2 to set Θ_0 is $\left|\theta^*_2 - \theta^*_{\{\frac{c}{1-c}\}}\right|$.

Let $\delta = \min\left(\frac{\left|\theta^*_1-\theta^*_{\{1-\frac{\varepsilon}{1-c}\}}\right|}{2}, \frac{\left|\theta^*_2-\theta^*_{\{\frac{c}{1-c}\}}\right|}{2}\right)$, when T is large enough, we have $\max\left(\left|\hat{\theta}_{\{1-\frac{\varepsilon}{1-c}\}} - \theta^*_{\{1-\frac{\varepsilon}{1-c}\}}\right|, \left|\hat{\theta}_{\{\frac{c}{1-c}\}} - \theta^*_{\{\frac{c}{1-c}\}}\right|\right) < \delta$ with probability $1-o_p(1)$.

$$\Theta_\delta = \{\theta : \theta \in (0,1] \text{ and } \exists\theta^*_\alpha \in \Theta_0 \text{ that } |\theta - \theta^*_\alpha| < \delta\}.$$

It's easy to figure out that $\theta^*_1, \theta^*_2 \notin \Theta_\delta$ but $\hat{\theta}_{\{1-\frac{\varepsilon}{1-c}\}}, \hat{\theta}_{\{\frac{c}{1-c}\}} \in \Theta_\delta$. Back to Equation (S5.62) and $\hat{\theta}_{1,\hat{\tau}} = \arg\max_{\theta\in(0,1+\epsilon)} \frac{\log L_2(\theta|\{G(t)\}_{t=0}^{\hat{\tau}})}{\hat{\tau}}$, when T is large enough, $\hat{\theta}_{1,\hat{\tau}}$ is closer to θ^*_2 than $\hat{\theta}_{\{1-\frac{\varepsilon}{1-c}\}}$ and closer to θ^*_1 than $\hat{\theta}_{\{\frac{c}{1-c}\}}$ with probability $1-o_p(1)$. So we get $\hat{\theta}_{1,\hat{\tau}} \in \Theta_\delta$ with probability $1-o_p(1)$.

By Theorem 4 in the main paper, $\hat{\theta}_{1,\tau^*} \xrightarrow{a.s.} \theta^*_1$, $\hat{\theta}_{2,\tau^*} \xrightarrow{a.s.} \theta^*_2$ and $\hat{\theta}_{2,\hat{\tau}} \xrightarrow{a.s.} \theta^*_2$ as T tends to infinity. We get the following inequalities with probability

$1 - o_p(1)$ when T is large enough for $\forall \delta_1 > 0$:

$$\begin{aligned}
&\left| \frac{\log L_2(\hat{\theta}_{1,\tau^*} | \{G(t)\}_{t=0}^{\tau^*})}{\tau^*} - E_{\theta_1^*}[\log(P(Y;\theta_1^*))] \right| \leq \delta_1, \\
&\left| \frac{\log L_2(\hat{\theta}_{1,\hat{\tau}} | \{G(t)\}_{t=0}^{\tau^*})}{\tau^*} - E_{\theta_1^*}[\log(P(Y;\hat{\theta}_{1,\hat{\tau}}))] \right| \leq \delta_1, \\
&\left| \frac{\log L_2(\hat{\theta}_{2,\tau^*} | \{G(t)\}_{t=\tau^*}^{\hat{\tau}})}{\hat{\tau} - \tau^*} - E_{\theta_2^*}[\log(P(Y;\theta_2^*))] \right| \leq \delta_1, \\
&\left| \frac{\log L_2(\hat{\theta}_{1,\hat{\tau}} | \{G(t)\}_{t=\tau^*}^{\hat{\tau}})}{\hat{\tau} - \tau^*} - E_{\theta_2^*}[\log(P(Y;\hat{\theta}_{1,\hat{\tau}}))] \right| \leq \delta_1, \\
&\left| \frac{\log L_2(\hat{\theta}_{2,\tau^*} | \{G(t)\}_{t=\hat{\tau}}^{T})}{T - \hat{\tau}} - E_{\theta_2^*}[\log(P(Y;\theta_2^*))] \right| \leq \frac{\delta_1}{2}, \\
&\left| \frac{\log L_2(\hat{\theta}_{2,\hat{\tau}} | \{G(t)\}_{t=\hat{\tau}}^{T})}{T - \hat{\tau}} - E_{\theta_2^*}[\log(P(Y;\theta_2^*))] \right| \leq \frac{\delta_1}{2}. \qquad \text{(S5.63)}
\end{aligned}$$

Let $\delta_1 = \min\left(\frac{E_{\theta_2^*}[\log(P(Y;\theta_2^*))] - \max_{\theta \in \Theta_\delta} E_{\theta_2^*}[\log(P(Y;\theta))]}{4}, \frac{E_{\theta_1^*}[\log(P(Y;\theta_1^*))] - \max_{\theta \in \Theta_\delta} E_{\theta_1^*}[\log(P(Y;\theta))]}{4} \right)$.

Equation (S5.63) implies:

$$\begin{aligned}
&\frac{\log L_2(\hat{\theta}_{1,\tau^*} | \{G(t)\}_{t=0}^{\tau^*})}{\tau^*} - \frac{\log L_2(\hat{\theta}_{1,\hat{\tau}} | \{G(t)\}_{t=0}^{\tau^*})}{\tau^*} \\
&\geq \frac{E_{\theta_1^*}[\log(P(Y;\theta_1^*))] - \max_{\theta \in \Theta_\delta} E_{\theta_1^*}[\log(P(Y;\theta))]}{2}, \\
&\frac{\log L_2(\hat{\theta}_{2,\tau^*} | \{G(t)\}_{t=\tau^*}^{\hat{\tau}})}{\hat{\tau} - \tau^*} - \frac{\log L_2(\hat{\theta}_{1,\hat{\tau}} | \{G(t)\}_{t=\tau^*}^{\hat{\tau}})}{\hat{\tau} - \tau^*} \\
&\geq \frac{E_{\theta_2^*}[\log(P(Y;\theta_2^*))] - \max_{\theta \in \Theta_\delta} E_{\theta_2^*}[\log(P(Y;\theta))]}{2}, \\
&\left| \frac{\log L_2(\hat{\theta}_{2,\tau^*} | \{G(t)\}_{t=\hat{\tau}}^{T})}{T - \hat{\tau}} - \frac{\log L_2(\hat{\theta}_{2,\hat{\tau}} | \{G(t)\}_{t=\hat{\tau}}^{T})}{T - \hat{\tau}} \right| \leq \delta_1. \quad \text{(S5.64)}
\end{aligned}$$

$\hat{\tau} - \tau^* > 0$ means that

$$\log L_2(\hat{\theta}_{1,\hat{\tau}} | \{G(t)\}_{t=0}^{\hat{\tau}}) + \log L_2(\hat{\theta}_{2,\hat{\tau}} | \{G(t)\}_{t=\hat{\tau}}^{T}) \geq \log L_2(\hat{\theta}_{1,\tau^*} | \{G(t)\}_{t=0}^{\tau^*}) + \log L_2(\hat{\theta}_{2,\tau^*} | \{G(t)\}_{t=\tau^*}^{T}).$$

Let

$$A = \quad \{[\log L_2(\hat{\theta}_{1,\hat{\tau}}|\{G(t)\}_{t=0}^{\hat{\tau}}) + \log L_2(\hat{\theta}_{2,\hat{\tau}}|\{G(t)\}_{t=\hat{\tau}}^{T}]$$

$$-[\log L_2(\hat{\theta}_{1,\tau^*}|\{G(t)\}_{t=0}^{\tau^*}) + \log L_2(\hat{\theta}_{2,\tau^*}|\{G(t)\}_{t=\tau^*}^{T})] \geq 0\}.$$

Further, for $\forall \varepsilon > 0,\ P\left(\hat{\tau} - \tau^* > \varepsilon T\right) = P\left(A, \hat{\tau} - \tau^* > \varepsilon T\right).$

Equation (S5.64) implies that $\left[\log L_2(\hat{\theta}_{1,\hat{\tau}}|\{G(t)\}_{t=1}^{\hat{\tau}}) + \log L_2(\hat{\theta}_{2,\hat{\tau}}|\{G(t)\}_{t=\hat{\tau}}^{T}\right] < \left[\log L_2(\hat{\theta}_{1,\tau^*}|\{G(t)\}_{t=1}^{\tau^*}) + \log L_2(\hat{\theta}_{2,\tau^*}|\{G(t)\}_{t=\tau^*}^{T}\right]$ with probability $1 - o_p(1)$ when T is large enough. It follows that

$$\lim_{T\to\infty} P\left(\hat{\tau} - \tau^* > \varepsilon T\right) = 0.$$

Consider $\tau^* - \hat{\tau} > \varepsilon T$. Divide the time range into three intervals $[0, T] = [0, \hat{\tau}] \cup [\hat{\tau} + 1, \tau^*] \cup [\tau^* + 1, T]$.

The remaining steps are similar to those for $\hat{\tau} - \tau^* > \varepsilon T$, when T is large enough, we have

$$\lim_{T\to\infty} P\left(\tau^* - \hat{\tau} > \varepsilon T\right) = 0.$$

Finally, it follows that

$$\lim_{T\to\infty} P\left(|\hat{\tau} - \tau| > \varepsilon T\right) = 0.$$

□

Proof of Theorem S.1. Let $\tilde{g}_i$ be the observed group label of node i. Under the null hypothesis with $\theta^* = 1$, we have:

$\tilde{D}_k(t)$ is the total degrees of the nodes with observed group label k at time t and $\tilde{Z}_k(t) = \tilde{D}_k(t) - \tilde{D}_k(t-1)$ is the increased degree of the observed group k at time t. We define the $\sigma-$algebra $\mathcal{F}_t = \sigma(\tilde{G}(0), \cdots, \tilde{G}(t))$, where $\tilde{G}(t)$ is the graph with the mislabeling at time t. Since $\theta^* = 1$, the probability of edges between nodes depends only on their degree. Let $\tilde{P}_k(t) = \frac{\tilde{D}_k(t-1)}{2(t-1)+n_0}$, $\tilde{p}_k = p_k - \sum_{k_1 \neq k} p_{k,k_1} + \sum_{k_1 \neq k} p_{k_1,k}$, we have

$$\begin{aligned} E[\tilde{Z}_k(t)|\mathcal{F}_{t-1}] = \quad & 2q\tilde{p}_k\tilde{P}_k(t) + q(1-\tilde{p}_k)\tilde{P}_k(t) + q\tilde{p}_k(1-\tilde{P}_k(t)) \\ & +2(1-q)\tilde{P}_k^2(t) + 2(1-q)(1-\tilde{P}_k(t))\tilde{P}_k(t) \\ = \quad & q\tilde{p}_k + (2-q)\tilde{P}_k(t). \end{aligned} \tag{S5.65}$$

By the same steps of Equations (S5.3)–(S5.7), let $\tilde{Z}_k^0(i) = \tilde{Z}_k(i) - E[\tilde{Z}_k(i)|\mathcal{F}_{i-1}]$, $\tilde{S}_k(t) = \sum_{i=1}^t \tilde{Z}_k^0(i)$.

$$\begin{aligned} \frac{\tilde{D}_k(t)}{2t+n_0} = \quad & \frac{\sum_{i=1}^t E[\tilde{Z}_k(i)|\mathcal{F}_{i-1}]}{2t} + o(1) \\ = \quad & \frac{q\tilde{p}_k}{2} + \frac{2-q}{2t}\sum_{i=1}^t \frac{\tilde{D}_k(i-1)}{2(i-1)+n_0} + o(1). \end{aligned} \tag{S5.66}$$

According to the Equation (S5.66), let:

$$f_k(x) = \frac{q\tilde{p}_k}{2} + \frac{2-q}{2}x. \tag{S5.67}$$

By the Banach fixed point theorem, $f_k(x) : (R, |\cdot|) \to (R, |\cdot|)$ is a contraction mapping with only one fixed point $x = \tilde{p}_k$.

By Equation (S5.66), when t is large enough, we have:

$$\left|\frac{\tilde{D}_k(t)}{2t+n_0} - \tilde{p}_k\right| = \frac{2-q}{2}\left|\frac{\sum_{i=1}^{t}\left(\frac{\tilde{D}_k(i-1)}{2(i-1)+n_0} - \tilde{p}_k\right)}{t}\right|, \tag{S5.68}$$

where $0 < (2-q)/2 < 1$.

Equation (S5.68) implies that $\frac{\tilde{D}_k(t)}{2t+n_0}$ approaches $\tilde{p}_k$ as $t \to \infty$. By Equations (S5.66)–(S5.68), we can deduce $\frac{\tilde{D}_k(t)}{2t} \xrightarrow{a.s.} \tilde{p}_k$.

Consider the likelihood function for θ of $\{\tilde{G}(t)\}_{t=0}^{T}$,

$$\begin{aligned}
\log L(\theta|\{\tilde{G}(t)\}_{t=0}^{T}) = \quad & \sum_{k=1}^{K}\sum_{t=1}^{T} \mathbb{1}\{v(t)=1, \tilde{g}_{e_1(t)} = \tilde{g}_{e_2(t)} = k\} \log[\tilde{P}_k(t) + (1-\theta)(1-\tilde{P}_k(t))] \\
& + \sum_{k=1}^{K}\sum_{t=1}^{T} \mathbb{1}\{v(t)=1, g_{e_1(t)} = k, g_{e_2(t)} \neq k\} \log[(1-\tilde{P}_k(t))\theta] \\
& + \sum_{k=1}^{K}\sum_{t=1}^{T} \mathbb{1}\{v(t)=0, \tilde{g}_{e_1(t)} = \tilde{g}_{e_2(t)} = k\} \log[\tilde{P}_k(t)[\tilde{P}_k(t) + (1-\theta)(1-\tilde{P}_k(t))]] \\
& + \sum_{k=1}^{K}\sum_{t=1}^{T} \mathbb{1}\{v(t)=0, \tilde{g}_{e_1(t)} = k, \tilde{g}_{e_2(t)} \neq k\} \log[\tilde{P}_k(t)(1-\tilde{P}_k(t))\theta].
\end{aligned} \tag{S5.69}$$

Repeat the same operations in Equation (S5.39). The score function for θ,

$$\begin{aligned}
& \frac{\partial}{\partial\theta}\log L(\theta|\{\tilde{G}(t)\}_{t=0}^{T}) \\
= \quad & -\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{v(t)=1,\tilde{g}_{e_1(t)}=\tilde{g}_{e_2(t)}=k\}(1-\tilde{P}_k(t))}{\tilde{P}_k(t)+(1-\theta)\left(1-\tilde{P}_k(t)\right)} \\
& -\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{v(t)=0,\tilde{g}_{e_1(t)}=\tilde{g}_{e_2(t)}=k\}(1-\tilde{P}_k(t))}{\tilde{P}_k(t)+(1-\theta)\left(1-\tilde{P}_k(t)\right)} \\
& +\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{v(t)=1,\tilde{g}_{e_1(t)}=k,\tilde{g}_{e_2(t)}\neq k\}}{\theta} \\
& +\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{v(t)=0,\tilde{g}_{e_1(t)}=k,\tilde{g}_{e_2(t)}\neq k\}}{\theta}. \qquad \text{(S5.70)}
\end{aligned}$$

Equation (S5.70) is a strictly decreasing continuous function of $\theta \in (0, 1+\epsilon)$, and Equation (S5.70) $\to \infty$ when $\theta \to 0$. Thus, there is a unique maximum point for $\log L(\theta|\{\tilde{G}(t)\}_{t=0}^{T})$ of $\theta \in (0, 1+\epsilon)$, where $\epsilon = \min\{\min\{\tilde{P}_k(t)/(1-\tilde{P}_k(t)) : t \in [1,T], \tilde{g}_{e_1(t)} = \tilde{g}_{e_2(t)} = k\} : k \in [K]\}$.

Let $\hat{\theta} = \arg\max_{\theta\in(0,1+\epsilon)} \log L(\theta|\{\tilde{G}(t)\}_{t=0}^{T})$. $\hat{\theta}$ is the unique maximum point for $\log L(\theta|\{\tilde{G}(t)\}_{t=0}^{T})$. So $\forall \theta \neq \hat{\theta} \in (0, 1+\epsilon)$, $\log L(\hat{\theta}|\{\tilde{G}(t)\}_{t=0}^{T}) > \log L(\theta|\{\tilde{G}(t)\}_{t=0}^{T})$. And it implies

$$\frac{1}{T}\log L(\hat{\theta}|\{\tilde{G}(t)\}_{t=0}^{T}) > \frac{1}{T}\log L(\theta|\{\tilde{G}(t)\}_{t=0}^{T}). \qquad \text{(S5.71)}$$

Same to the Equation (S5.41), by LLN and Theorem 1 in the main paper, get

$$\begin{aligned} & \lim_{T\to\infty} \frac{\log L(\theta|\{\tilde{G}_t\}_{t=0}^T)}{T} \\ = & \sum_{k=1}^{K} \left[\log(\tilde{p}_k + (1-\theta)(1-\tilde{p}_k)) \times \tilde{p}_k^2 + \log\left((1-\tilde{p}_k)\,\theta\right) \times \tilde{p}_k(1-\tilde{p}_k)\right]. \end{aligned} \tag{S5.72}$$

Consider a random variable $Y = (x, z)$ such that $x \in \{1, \cdots, K\}$ and $z \in \{0, 1\}$. The probability distribution of Y is

$$P(y;\theta) = \begin{cases} \tilde{p}_k(1-\tilde{p}_k)\theta, & \text{if } x = k, y = 0; \\ \tilde{p}_k^2 + \tilde{p}_k(1-\tilde{p}_k)(1-\theta), & \text{if } x = k, y = 1. \end{cases} \tag{S5.73}$$

For any other $\theta \in (0, 1+\epsilon^*)$, where $\epsilon^* = \min_{k\in[K]}\{\tilde{p}_k/(1-\tilde{p}_k)\}$, for $\theta^* = 1$

$$\begin{aligned} E_1[\log[P(Y;1)/P(Y;\theta)]] = & \quad -E_1[\log[P(Y;\theta)/P(Y;1)]] \\ \geq & \quad -\log[E_1(P(Y;\theta)/P(Y;1))] \\ = & \quad -\log[1] = 0. \end{aligned} \tag{S5.74}$$

Equation (S5.74) implies that

$$\forall\theta \in (0, 1+\epsilon^*),\ E_1[\log(P(Y;1))] > E_1[\log(P(Y;\theta))]. \tag{S5.75}$$

Moreover, we can prove $E_1[\log(P(Y;1))] = \lim_{T\to\infty} \frac{\log L(1|\{\tilde{G}(t)\}_{t=0}^T)}{T} + \sum_{k=1}^K \tilde{p}_k \log(\tilde{p}_k)$. Thus, by Equation (S5.75), when T is large enough, we know $\frac{\log L(1|\{\tilde{G}(t)\}_{t=0}^T)}{T} > \frac{\log L(\theta|\{\tilde{G}(t)\}_{t=0}^T)}{T}$, $\forall \theta \in (0, 1+\epsilon^*)$. Combining this result with Equation (S5.71), get $P(\lim_{T\to\infty}\hat{\theta} = 1) = 1$, which means $\lim_{T\to\infty}\hat{\theta} \xrightarrow{a.s.} 1$.

Let $l(\theta|\{\tilde{G}(t)\}_{t=0}^T) = \frac{\partial \log L(\theta|\{\tilde{G}(t)\}_{t=0}^T)}{\partial\theta}$. By Equation (S5.70), we get:

$$\begin{aligned} l'(\theta|\{\tilde{G}(t)\}_{t=0}^T) = & \quad \frac{\partial l'(\theta|\{\tilde{G}(t)\}_{t=0}^T)}{\partial\theta} \\ = & \quad -\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{\tilde{g}_{e_1(t)} = \tilde{g}_{e_2(t)} = k\}\left(1-\tilde{P}_k(t)\right)^2}{\left[\tilde{P}_k(t) + (1-\theta)\left(1-\tilde{P}_k(t)\right)\right]^2} \\ & \quad -\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{\tilde{g}_{e_1(t)} = k, \tilde{g}_{e_2(t)} \neq k\}}{\theta^2}. \end{aligned}$$

Taking the Taylor Expansion of $l(\theta|\{\tilde{G}(t)\}_{t=0}^T)$ around 1, we can get:

$$l(\theta|\{\tilde{G}(t)\}_{t=0}^T) = l(1|\{\tilde{G}(t)\}_{t=0}^T) + (\theta-1)l'(1|\{\tilde{G}(t)\}_{t=0}^T) + (\theta-1)^2 l'(\theta^{**}|\{\tilde{G}(t)\}_{t=0}^T), \tag{S5.76}$$

where $\theta^{**} = 1 + \xi \circ (\theta - 1)$, $\xi \in [0,1]$. $l(\hat{\theta}|\{G_t\}_{t=0}^T) = 0$ when T is large enough. Equation (S5.76) implies

$$T^{1/2}(\hat{\theta}-1) = -\frac{\frac{T^{1/2}l(1|\{\tilde{G}(t)\}_{t=0}^T)}{T}}{\frac{l'(1|\{\tilde{G}(t)\}_{t=0}^T)}{T} + \frac{l'(\theta^{**}|\{\tilde{G}(t)\}_{t=0}^T)(\hat{\theta}-1)}{T}}. \tag{S5.77}$$

Consider the random variable Y and its probability distribution $P(\cdot;\theta^*)$

defined by Equation (S5.73).

$$
\begin{aligned}
I(1) = \quad & E_1\left[-\frac{\partial^2 \log P(Y;\theta)}{\partial \theta^2}\Big|_{\theta=1}\right] = E_1\left[\left(\frac{P'(Y;1)}{P(Y;1)}\right)^2\right] \\
= \quad & \sum_y \frac{P'(y;1)^2}{P(y;1)} = \sum_{k=1}^{K}\left[\tilde{p}_k(1-\tilde{p}_k) + (1-\tilde{p}_k)^2\right] = K-1.
\end{aligned}
\tag{S5.78}
$$

By the martingale convergence theorem, martingale difference central limit theorem and Equation (S5.78), we have that when T tends to infinity,

$$
\hat{\theta} \xrightarrow{a.s.} 1, \quad T^{1/2}\frac{l(1|\{\tilde{G}(t)\}_{t=0}^{T})}{T} \xrightarrow{d} N(0, I(1)), \tag{S5.79}
$$

$$
-\frac{l'(1|\{\tilde{G}(t)\}_{t=0}^{T})}{T} \xrightarrow{a.s.} I(1), \tag{S5.80}
$$

and

$$
-\frac{l'(\theta^{**}|\{\tilde{G}(t)\}_{t=0}^{T})}{T} \xrightarrow{a.s.} \sum_{k=1}^{K}\left[\frac{\tilde{p}_k(1-\tilde{p}_k)}{(\theta^{**})^2} + \frac{\tilde{p}_k^2(1-\tilde{p}_k)^2}{(\tilde{p}_k + (1-\tilde{p}_k)(1-\theta^{**}))^2}\right]. \tag{S5.81}
$$

By Equations (S5.79)–(S5.81), it follows that when T tends to infinity,

$$
T^{1/2}(\hat{\theta}-1) \xrightarrow{d} N(0, \frac{1}{K-1}).
$$

□

Proof of Theorem S.2. Set t_i be the time of node i added to the network. Consider a node w in node u's childset $\mathcal{C}_u$, let $g_u = k$. The probability of node w connected to group k ($\mathcal{O}_{w,k}$) conditional on the g_w is

$$\begin{aligned} P(\mathcal{O}_{w,k}|g_w = k) = & \quad \frac{D_k(t_w)}{2t_w + n_0} + (1-\theta)\left(1 - \frac{D_k(t_w)}{2t_w + n_0}\right); \\ P(\mathcal{O}_{w,k}|g_w \neq k) = & \quad \theta \frac{D_k(t_w)}{2t_w + n_0}. \end{aligned} \tag{S5.82}$$

Given g_u, we get:

$$P(g_w|\mathcal{O}_{w,k}) = \quad \frac{p_{g_w} P(\mathcal{O}_{w,k}|g_w)}{\sum_{g_w=1}^{K} p_{g_w} P(\mathcal{O}_{w,k}|g_w)}.$$

Let $n_u = |\mathcal{C}_u|$, $n_{u,j} = \sum_{w \in \mathcal{C}_u} \mathbb{1}\{g_w = j\}$ for $j \in \{1, \cdots, K\}$. By LLN,

$$\frac{n_{u,j}}{p_j n_u} - E\left(\frac{n_{u,j}}{p_j n_u}\right) \xrightarrow{i.p.} 0.$$

By Equation (S5.82), for $\theta \neq 1$ and for each $j \neq k$,

$$E\left(\frac{n_{u,j}}{p_j n_u}\right) < E\left(\frac{n_{u,k}}{p_k n_u}\right).$$

It implies that

$$\hat{g}_u = \arg\max_j \frac{n_{u,j}}{p_j n_u} \xrightarrow{i.p.} k.$$

□

Proof of Theorem S.3. Assume parameter q changed at time $\{\tau_l\}_{l=1}^{L}$, where for each l, $0 < \tau_l < T$, $\tau_l/T = c_l \in (0,1)$. Set

$$q = \begin{cases} q_1, & \text{if } t \le \tau_1; \\ q_2, & \text{if } \tau_1 < t \le \tau_2; \\ \vdots & \\ q_{L+1}, & \text{if } \tau_L < t \le T. \end{cases}$$

Further, let $\tau_0 = 0$, $\tau_{L+1} = T$; $c_0 = 0$, $c_{L+1} = 1$.

Let $Z_k(t) = D_k(t) - D_k(t-1), t \ge 1$ be the increased degree of group k at time t. We define the $\sigma-$algebra $\mathcal{F}_t = \sigma(G(0), \cdots, G(t))$. Let $P_k(t) = \frac{D_k(t-1)}{2(t-1)+n_0}$, by Equation (S5.2) , we have

$$E(Z_k(t)|\mathcal{F}_{t-1}) = \begin{cases} (2-\theta)q_1 p_k + [2 - q_1(2-\theta)]P_k(t), & \text{if } t \le \tau_1; \\ (2-\theta)q_2 p_k + [2 - q_2(2-\theta)]P_k(t), & \text{if } \tau_1 < t \le \tau_2; \\ \vdots & \\ (2-\theta)q_{L+1} p_k + [2 - q_{L+1}(2-\theta)]P_k(t), & \text{if } t > \tau_L. \end{cases}$$

According to Equations (S5.3)–(S5.8), when t is large enough,

$$\frac{D_k(t)}{2t+n_0} = \quad \frac{\sum_{i=1}^{t} E(Z_k(i)|\mathcal{F}_{i-1})}{2t} + o(1).$$

Thus, for $\{\tau_1, \cdots, \tau_L, T\}$ we can get:

$$\begin{cases} \frac{D_k(\tau_1)}{2\tau_1+n_0} = \frac{(2-\theta)p_k}{2} q_1 + \frac{[2-q_1(2-\theta)]}{2\tau_1} \sum_{t=1}^{\tau_1} \frac{D_k(t-1)}{2(t-1)+n_0} + o(1) \\ \frac{D_k(\tau_2)}{2\tau_2+n_0} = \frac{(2-\theta)p_k}{2} \frac{c_1q_1+(c_2-c_1)q_2}{c_2} + \frac{[2-q_1(2-\theta)]}{2\tau_2} \sum_{t=1}^{\tau_1} \frac{D_k(t-1)}{2(t-1)+n_0} + \frac{[2-q_2(2-\theta)]}{2\tau_2} \sum_{t=\tau_1+1}^{\tau_2} \frac{D_k(t-1)}{2(t-1)+n_0} + o(1) \\ \vdots \\ \frac{D_k(T)}{2T+n_0} = \frac{(2-\theta)p_k}{2} (\sum_{l=1}^{L+1} (c_l - c_{l-1})q_l) + \sum_{l=1}^{L+1} \frac{[2-q_l(2-\theta)]}{2T} \sum_{t=1+\tau_{l-1}}^{\tau_l} \frac{D_k(t-1)}{2(t-1)+n_0} + o(1) \end{cases}$$

$\frac{\sum_{i=1}^{l}(c_i-c_{i-1})q_i}{c_l} \in (0,1)$ for $l = 1, 2, \cdots, L+1$,

$$\begin{cases} \left|\frac{D_k(\tau_1)}{2\tau_1+n_0} - p_k\right| = \frac{[2-q_1(2-\theta)]}{2\tau_1} \left|\sum_{t=1}^{\tau_1} \frac{D_k(t-1)}{2(t-1)+n_0} - p_k\right| + o(1) \\ \left|\frac{D_k(\tau_2)}{2\tau_2+n_0} - p_k\right| = \frac{c_1}{c_2} \frac{[2-q_1(2-\theta)]}{2\tau_1} \left|\sum_{t=1}^{\tau_1} \frac{D_k(t-1)}{2(t-1)+n_0} - p_k\right| + \frac{c_2-c_1}{c_2} \frac{[2-q_2(2-\theta)]}{2(\tau_2-\tau_1)} \left|\sum_{t=\tau_1+1}^{\tau_2} \frac{D_k(t-1)}{2(t-1)+n_0} - p_k\right| + o(1) \\ \vdots \\ \left|\frac{D_k(T)}{2T+n_0} - p_k\right| = \sum_{l=1}^{L+1} (c_l - c_{l-1}) \frac{[2-q_l(2-\theta)]}{2(\tau_l-\tau_{l-1})} \left|\sum_{t=1+\tau_{l-1}}^{\tau_l} \frac{D_k(t-1)}{2(t-1)+n_0} - p_k\right| + o(1) \end{cases}$$

where $0 < |2 - q_l(2-\theta)|/2 < 1$ for $l = 1, \cdots, L+1$.

Finally, we can deduce $\frac{D_k(\tau_l)}{2\tau_l} \xrightarrow{a.s.} p_k$ $l = 1, \cdots, L+1$.

□

Proof of Theorem S.4. The log likelihood function for the KPA model with the parameter p underwent a finite number of changesis:

$$\sum_{l=1}^{L+1} \log(\psi_l | \{G(t)\}_{t=\tau_{l-1}}^{\tau_l}),$$

where $\psi_l = (\theta, \{p_k\}_{k=1}^{K}, q_l)$, and $\log(\psi|\{G(t)\})$ is defined in Equation (S5.36) in the supplementary.

Considering the score function for θ, we have

$$\begin{aligned}
&\frac{\partial}{\partial\theta}\sum_{l=1}^{L+1}\log(\psi_l|\{G(t)\}_{t=\tau_{l-1}}^{\tau_l}) \\
= \quad & -\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{v(t)=1, g_{e_1(t)}=g_{e_2(t)}=k\}(1-P_k(t))}{P_k(t)+(1-\theta)(1-P_k(t))} \\
& -\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{v(t)=0, g_{e_1(t)}=g_{e_2(t)}=k\}(1-P_k(t))}{P_k(t)+(1-\theta)(1-P_k(t))} \\
& +\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{v(t)=1, g_{e_1(t)}=k, g_{e_2(t)}\neq k\}}{\theta} \\
& +\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\mathbb{1}\{v(t)=0, g_{e_1(t)}=k, g_{e_2(t)}\neq k\}}{\theta}. \qquad \text{(S5.83)}
\end{aligned}$$

Equation (S5.83) is a strictly decreasing continuous function of $\theta \in (0, 1+\epsilon)$, where $\epsilon = \min\{\min\{P_k(t)/(1-P_k(t)) : t \in [1,T], g_{e_1(t)} = g_{e_2(t)} = k\} : k \in [K]\}$. Equation (S5.83) $\to \infty$ when $\theta \to 0$ and Equation (S5.83) $\to -\infty$ when $\theta \to 1+\epsilon$. Thus, there is a unique root for $\frac{\partial}{\partial\theta}\sum_{l=1}^{L+1}\log(\psi_l|\{G(t)\}_{t=\tau_{l-1}}^{\tau_l})$ of $\theta \in (0, 1+\epsilon)$.

Let $\log L_2(\theta|\{G(t)\}_{t=0}^{T}) = \frac{\partial}{\partial\theta}\sum_{l=1}^{L+1}\log(\psi_l|\{G(t)\}_{t=\tau_{l-1}}^{\tau_l})$, $\hat{\theta} = \arg\max_{\theta\in(0,1+\epsilon)}\log L_2(\theta|\{G(t)\}_{t=0}^{T})$. $\hat{\theta}$ is the unique maximum point for $\log L_2(\theta|\{G(t)\}_{t=0}^{T})$. Thus, $\forall\theta \neq \hat{\theta} \in (0, 1+\epsilon)$, $\log L_2(\hat{\theta}|\{G(t)\}_{t=0}^{T}) > \log L_2(\theta|\{G(t)\}_{t=0}^{T})$. And it implies

$$\frac{1}{T}\log L_2(\hat{\theta}|\{G(t)\}_{t=0}^{T}) > \frac{1}{T}\log L_2(\theta|\{G(t)\}_{t=0}^{T}). \qquad \text{(S5.84)}$$

Let $\log L_2(\theta|\{G(t)\}_{t=0}^{T}) = \sum_{l=1}^{L+1}\log L_2(\theta|\{G(t)\}_{t=\tau_{l-1}}^{\tau_l})$, $\log L_2(\theta|\{G(t)\}_{t=\tau_{l-1}}^{\tau_l}) = \frac{\partial}{\partial\theta}\log(\psi_l|\{G(t)\}_{t=\tau_{l-1}}^{\tau_l})$.

By LLN and Theorem S.3,

$$
\begin{aligned}
& \lim_{T\to\infty} \frac{\log L_2(\theta|\{G(t)\}_{t=\tau_{l-1}}^{\tau_l})}{\tau_l - \tau_{l-1}} \\
= \quad & \sum_{k=1}^{K} \log(p_k^* + (1-\theta)(1-p_k^*)) \times [(p_k^*)^2 + p_k^*(1-p_k^*)(1-\theta^*)] \\
& + \sum_{k=1}^{K} \log\left((1-p_k^*)\,\theta\right) \times p_k^*(1-p_k^*)\theta^*. \\
& \lim_{T\to\infty} \frac{\log L_2(\theta|\{G_t\}_{t=0}^{T})}{T} \\
= \quad & \sum_{k=1}^{K} \log(p_k^* + (1-\theta)(1-p_k^*)) \times [(p_k^*)^2 + p_k^*(1-p_k^*)(1-\theta^*)] \\
& + \sum_{k=1}^{K} \log\left((1-p_k^*)\,\theta\right) \times p_k^*(1-p_k^*)\theta^*.
\end{aligned}
$$

Consider a random variable $Y = (x, z)$ such that $x \in \{1, \cdots, K\}$ and $z \in \{0, 1\}$. The probability distribution of Y is

$$
P(y;\theta) = \begin{cases} p_k^*(1-p_k^*)\theta, & \text{if } x = k, y = 0; \\ (p_k^*)^2 + p_k^*(1-p_k^*)(1-\theta), & \text{if } x = k, y = 1. \end{cases}
$$

For any other $\theta \neq \theta^* \in (0, 1+\epsilon^*)$, where $\epsilon^* = \min_{k\in[K]}\{p_k^*/(1-p_k^*)\}_k$.

$$
\begin{aligned}
E_{\theta^*} \log[P(Y;\theta^*)/P(Y;\theta)] = & \quad -E_{\theta^*} \log[P(Y;\theta)/P(Y;\theta^*)] \\
\geq & \quad -\log[E_{\theta^*}(P(Y;\theta)/f(Y;\theta^*))] \\
= & \quad -\log[1] = 0. \qquad \text{(S5.85)}
\end{aligned}
$$

Equation (S5.85) implies that

$$
\forall \theta \neq \theta^* \in (0, 1+\epsilon^*),\ E_{\theta^*} \log[P(Y;\theta^*)] > E_{\theta^*} \log[P(Y;\theta)]. \qquad \text{(S5.86)}
$$

Moreover, we can prove $E_{\theta^*}\log[P(Y;\theta^*)] = \lim_{T\to\infty}\frac{\log L_2(\theta^*|\{G_t\}_{t=0}^T)}{T} + \sum_{k=1}^K p_k^*\log(p_k^*)$. Thus, by Equation (S5.86), when T is large enough, we know $\frac{\log L_2(\theta^*|\{G_t\}_{t=0}^T)}{T} > \frac{\log L_2(\theta|\{G_t\}_{t=0}^T)}{T}$, $\forall\theta\neq\theta^*\in(0,1+\epsilon^*)$. Combining this result with Equation (S5.84), get $P(\lim_{T\to\infty}\hat{\theta}=\theta^*)=1$, which means

$$\lim_{T\to\infty}\hat{\theta}\xrightarrow{a.s.}\theta^*.$$

Further, consider the snapshot estimation, the proof is the same to Theorem 6.

$$\begin{aligned}
\tilde{\theta} = &\quad \arg\min_{\theta\in(0,1+\tilde{\epsilon})} L_T(\theta|G(T)),\\
\frac{L_T(\theta|G_T)}{T} = &\quad \sum_{k=1}^K\left[\frac{|\mathcal{E}_{k,1}(T)|}{T}\log(\frac{D_k(T)}{2|\mathcal{E}(T)|}(\frac{D_k(T)}{2|\mathcal{E}(T)|}+(1-\theta)(1-\frac{D_k(T)}{2|\mathcal{E}(T)|})))\right]\\
&\quad +\sum_{k=1}^K\left[\frac{|\mathcal{E}_{k,0}(T)|}{T}\log(\frac{D_k(T)}{2|\mathcal{E}(T)|}\theta(1-\frac{D_k(T)}{2|\mathcal{E}(T)|}))\right].
\end{aligned}$$

By LLN and Theorem S.3, we have

$$\begin{aligned}
&\lim_{T\to\infty}\frac{L_T(\theta|G(T))}{T}\\
\xrightarrow{a.s.}\quad &\sum_{k=1}^K[((p_k^*)^2+p_k^*(1-p_k^*)(1-\theta^*))\log((p_k^*)^2+p_k^*(1-p_k^*)(1-\theta))\\
&+(p_k^*(1-p_k^*)\theta^*)\log(p_k^*(1-p_k^*)\theta)]\\
=\quad & E_{\theta^*}[\log P(Y;\theta)]. \qquad \text{(S5.87)}
\end{aligned}$$

Equation (S5.87) implies $\frac{1}{T}L_T(\theta^*|G(T)) > \frac{1}{T}L_T(\theta|G(T))$, $\forall\theta\neq\theta^*\in(0,1+\tilde{\epsilon})$ when T is large enough. We can get $\tilde{\theta}=\arg\max_{\theta\in(0,1+\tilde{\epsilon})}L_T(\theta|G(T))\xrightarrow{a.s.}\theta^*$, where $\tilde{\epsilon}=\min_{k\in[K]}\{\frac{D_k(T)}{2|\mathcal{E}(T)|}/(1-\frac{D_k(T)}{2|\mathcal{E}(T)|})\}_k$.

□

Bibliography

Chung, F. and L. Lu (2006). *Complex graphs and networks*. Number 107. American Mathematical Soc.

Hajek, B. and S. Sankagiri (2019). Community recovery in a preferential attachment graph. *IEEE Transactions on Information Theory 65*(11), 6853–6874.