Supplementary Material

A. Full Results for the Real Data Experiment

We record the full results for our real data experiment in Tables 3 and 4.

Data			LongSync			MultiSync-New			IRLS-New			MPLS on full dataset			Remaining
	n	k	$\bar{e}$	$\hat{e}$	t	$\bar{e}$	$\hat{e}$	t	$\bar{e}$	$\hat{e}$	t	$\bar{e}$	$\hat{e}$	t	cameras
Alamo	627	10	**2.67**	**1.03**	6.68	3.21	1.81	80.77	2.74	1.14	3.58	2.03	0.95	53.67	0.70
Ellis Island	247	7	**1.07**	**0.60**	1.05	1.81	1.37	24.46	1.29	0.92	0.79	0.89	0.48	4.40	0.66
Madrid Metropolis	394	6	**2.98**	**1.89**	1.28	4.42	3.92	14.76	3.35	2.27	1.00	2.10	1.10	5.76	0.53
Montreal Notre Dame	474	8	**3.89**	**0.45**	2.96	4.34	1.05	39.17	4.00	0.60	1.86	0.78	0.41	14.95	0.70
Notre Dame	553	11	**1.00**	**0.60**	5.26	1.93	1.64	111.98	1.40	1.12	3.04	0.96	0.50	52.68	0.66
NYC Library	376	6	**2.05**	**1.09**	1.20	2.56	1.63	14.57	2.31	1.47	0.83	1.56	1.01	5.02	0.56
Piazza del Popolo	345	7	**3.83**	**0.62**	1.39	4.07	1.07	22.91	3.89	0.83	0.98	1.61	0.57	6.26	0.61
Roman Forum	1102	6	**2.47**	**1.61**	5.60	2.84	1.98	16.13	2.55	1.68	3.21	1.80	1.33	24.05	0.48
Tower of London	489	5	**2.85**	**2.18**	1.82	3.49	2.66	8.43	2.94	2.20	1.17	2.50	2.13	5.63	0.60
Union Square	930	4	8.02	4.31	2.59	7.79	3.93	4.92	**7.76**	**3.73**	1.73	4.50	3.53	7.20	0.42
Vienna Cathedral	918	9	**3.65**	**0.58**	5.66	4.34	1.57	56.45	3.76	0.77	3.36	1.30	0.53	54.62	0.48
Gendarmenmarkt	742	6	84.95	**77.60**	3.21	83.30	80.48	15.75	**74.71**	84.23	2.27	48.52	40.16	13.58	0.47
Piccadilly	2508	9	**5.07**	**1.75**	22.56	5.43	2.46	65.89	5.21	2.08	10.72	2.20	1.45	429.40	0.45
Trafalgar	5433	9	**7.12**	**2.33**	79.33	10.01	5.19	91.69	7.25	2.44	41.59	1.88	1.17	1796.41	0.38
Yorkminster	458	6	**1.97**	**1.36**	2.52	2.33	1.74	14.97	2.01	1.38	1.89	1.63	1.31	7.05	0.61

Table 3. Results for PhotoTourism. For each dataset, $\bar{e}$ and $\hat{e}$ indicate the mean error and median error of the output absolute rotation estimates measured in degrees, and t is the total runtime of each method measured in seconds. The last column indicates the remaining portion of cameras for each dataset after adpoting our new graph preprocessing method.

Data			LongSync-Naive			MultiSync			IRLS			MPLS on full dataset		
	n	k	$\bar{e}$	$\hat{e}$	t	$\bar{e}$	$\hat{e}$	t	$\bar{e}$	$\hat{e}$	t	$\bar{e}$	$\hat{e}$	t
Alamo	627	10	**7.45**	**1.11**	8.91	7.74	1.56	81.73	7.56	1.30	5.71	3.67	1.02	55.43
Ellis Island	247	7	**3.84**	**0.69**	2.28	5.24	2.29	25.49	4.12	1.09	1.77	2.82	0.50	5.72
Madrid Metropolis	394	6	**9.85**	**2.92**	2.96	10.27	3.91	15.46	10.13	3.61	2.24	5.83	1.31	7.63
Montreal Notre Dame	474	8	**5.93**	**0.61**	5.17	6.49	1.44	41.20	6.06	0.87	3.37	1.13	0.50	18.40
Notre Dame	553	11	**4.57**	**0.72**	8.56	4.97	1.28	117.81	4.82	1.09	5.19	2.71	0.64	57.94
NYC Library	376	6	**6.15**	**1.65**	2.77	7.13	2.96	15.58	6.28	1.92	2.07	3.11	1.30	5.92
Piazza del Popolo	345	7	**6.37**	**0.99**	2.84	10.18	7.09	33.35	7.23	1.18	2.07	3.44	0.86	7.19
Roman Forum	1102	6	**5.98**	**1.80**	10.15	6.61	2.57	19.17	6.06	1.93	5.81	2.87	1.41	27.33
Tower of London	489	5	**6.46**	**2.95**	3.99	7.03	3.43	9.77	6.74	3.24	2.77	3.96	2.44	6.53
Union Square	930	4	25.68	**5.68**	5.95	27.64	7.24	7.34	**25.31**	5.74	4.20	6.14	3.70	8.52
Vienna Cathedral	918	9	**13.26**	**1.60**	10.31	13.74	2.37	62.84	13.47	1.97	6.33	6.19	1.31	58.05
Gendarmenmarkt	742	6	**74.25**	72.34	6.40	74.63	**71.53**	17.56	76.58	81.32	4.33	39.70	10.48	17.13
Piccadilly	2508	9	**9.58**	**2.82**	42.17	9.91	3.19	72.09	10.66	3.78	19.39	4.45	2.08	455.83
Trafalgar	5433	9	**9.61**	**3.27**	130.55	10.23	4.16	112.62	9.75	3.44	60.49	5.49	4.39	1929.21
Yorkminster	458	6	**8.25**	**1.69**	5.06	8.80	2.47	16.59	8.28	1.74	3.87	3.55	1.58	8.31

Table 4. Results for PhotoTourism where all methods are performed without our graph preprocessing method. For each dataset, $\bar{e}$ and $\hat{e}$ indicate the mean error and median error of the output absolute rotation estimates measured in degrees, and t is the total runtime of each method measured in seconds. The last column indicates the remaining portion of cameras for each dataset after the camera pruning step of our improved pipeline.

B. Proof for the Formulas of g_c and f_c and their Computation Complexity

In this section we prove the formulas and time complexity for f_c and g_c defined in section 3.

For $c=3$, since all 3-cycles are simple, $f_c(\boldsymbol{W})(i,j)=\sum_{L\in C_{ij}^c}\prod_{e\in L\setminus\{ij\}}=\sum_{k\in[n]}w_{ik}w_{kj}$ is exactly the ij-th entry of $\boldsymbol{W}^2$, and $g_c(\boldsymbol{W},\boldsymbol{R})(i,j)=\sum_{L\in C_{ij}^c}=\sum_{k\in[n]}w_{ik}\boldsymbol{R}_{ik}w_{kj}\boldsymbol{R}_{kj}$ is exactly the ij-th block of $\boldsymbol{P}^2$.

For $c\geq 4$, there are redundant cycles in C_{ij}^c, i.e. cycles that are not simple. We follow the argument in [35] to compute $f_c(\boldsymbol{W})(i,j)$ and $g_c(\boldsymbol{W},\boldsymbol{R})(i,j)$. For example, the cycle $ikij$ is redundant since the node i repeats twice. We say this cycle satisfy the partition $0+2+1$ of $c-1$, in that the number of steps from the first node to the repeated node is 0, the number of steps from the repeated node to its second appearance is 2, and the number of remaining steps to the last letter is 1. Some cycles may satisfy more than 1 partition. For integer $1\leq a\leq c-1$, let $C_{ij,a}^c$ be the set of redundant c-cycles satisfying a partitions. Let q_c be the number of admissible partitions of length c, i.e. partitions that correspond to a redundant cycle. Then the function f_c and g_c can be written as follows:

$$f_c(\boldsymbol{W})(i,j)=\boldsymbol{W}^{c-1}+\sum_{a=1}^{q_c}(-1)^a\sum_{L\in C_{ij,a}^c}\prod_{e\in L\setminus\{ij\}}w_e \tag{15}$$

$$g_c(\boldsymbol{W},\boldsymbol{R})(i,j)=\boldsymbol{P}^{c-1}+\sum_{a=1}^{q_c}(-1)^a\sum_{L\in C_{ij,a}^c}\prod_{e\in L\setminus\{ij\}}w_e\boldsymbol{R}_e. \tag{16}$$

For $c=4$, the set of admissible partitions is $\{0+2+1,1+2+0\}$, therefore $q_4=2$. By enumerating the possible cycles for any combination of such admissible partitions, we know that the set $C_{ij,1}^4=\{k\in[n]:ikij\}\cup\{k\in[n]:ijkj\}$, and the set $C_{ij,2}^4=\{ijij\}$. Therefore we can simplify the above formulation as:

$$f_c(\boldsymbol{W})(i,j)=\boldsymbol{W}^{c-1}-\sum_{k\in[n]}w_{ik}w_{ki}w_{ij}-\sum_{k\in[n]}w_{ij}w_{jk}w_{kj}+w_{ij}w_{ji}w_{ij} \tag{17}$$

$$g_c(\boldsymbol{W},\boldsymbol{R})(i,j)=\boldsymbol{P}^{c-1}-\sum_{k\in[n]}w_{ik}w_{ki}w_{ij}\boldsymbol{R}_{ik}\boldsymbol{R}_{ki}\boldsymbol{R}_{ij}-\sum_{k\in[n]}w_{ij}w_{jk}w_{kj}\boldsymbol{R}_{ij}\boldsymbol{R}_{jk}\boldsymbol{R}_{kj}+w_{ij}w_{ji}w_{ij}\boldsymbol{R}_{ij}\boldsymbol{R}_{ji}\boldsymbol{R}_{ij}. \tag{18}$$

This can be vectorized as

$$f_c(\boldsymbol{W})=\boldsymbol{W}^3-\mathrm{d}(\boldsymbol{W}^2)\boldsymbol{W}-\boldsymbol{W}\mathrm{d}(\boldsymbol{W}^2)+\boldsymbol{W}^{\odot 3} \tag{19}$$

$$g_c(\boldsymbol{W},\boldsymbol{R})=\boldsymbol{P}^3-\mathrm{d}(\boldsymbol{P}^2)\boldsymbol{P}-\boldsymbol{P}\mathrm{d}(\boldsymbol{P}^2)+\boldsymbol{P}^{\odot 3}. \tag{20}$$

Using similar arguments as above (one may refer to [35]), we have the formulas for $c=5$ and $c=6$. The formulas for $c=5$ are presented in Table 1. The formulas for $c=6$ are as follows:

$$\begin{aligned}
f_c(\boldsymbol{W})=&\boldsymbol{W}\mathrm{d}(\boldsymbol{W}^4)+\mathrm{d}(\boldsymbol{W}^4)\boldsymbol{W}+\boldsymbol{W}^2\mathrm{d}(\boldsymbol{W}^3)+\mathrm{d}(\boldsymbol{W}^3)\boldsymbol{W}^2+\boldsymbol{W}\mathrm{d}(\boldsymbol{W}^2)\boldsymbol{W}^2+\boldsymbol{W}^2\mathrm{d}(\boldsymbol{W}^2)\boldsymbol{W}+\boldsymbol{W}\mathrm{d}(\boldsymbol{W}^3)\boldsymbol{W}\\
&+\boldsymbol{W}^2\odot\boldsymbol{W}^{\odot 3}+3\boldsymbol{W}\odot(\boldsymbol{W}^{\odot 2})^2+2\boldsymbol{W}\mathrm{d}(\boldsymbol{W}^2)\odot\boldsymbol{W}^{\odot 2}+2\mathrm{d}(\boldsymbol{W}^2)\boldsymbol{W}\odot\boldsymbol{W}^{\odot 2}\\
&+4\mathrm{d}(\boldsymbol{W}^2)\boldsymbol{W}^{\odot 3}+4\boldsymbol{W}^{\odot 3}\mathrm{d}(\boldsymbol{W}^2)-\boldsymbol{W}\mathrm{d}(\boldsymbol{W}\mathrm{d}(\boldsymbol{W}^2)\boldsymbol{W})-\mathrm{d}(\boldsymbol{W}\mathrm{d}(\boldsymbol{W}^2)\boldsymbol{W})\boldsymbol{W}\\
&-2\boldsymbol{W}(\boldsymbol{W}^{\odot 2}\odot\boldsymbol{W}^2)-2(\boldsymbol{W}^{\odot 2}\odot\boldsymbol{W}^2)\boldsymbol{W}-\boldsymbol{W}^{\odot 2}\boldsymbol{W}^2-\boldsymbol{W}^2\boldsymbol{W}^{\odot 2}\\
&-2\boldsymbol{W}\mathrm{d}(\boldsymbol{W}^2)^2-2\mathrm{d}(\boldsymbol{W}^2)^2\boldsymbol{W}-\boldsymbol{W}(\boldsymbol{W}\odot\boldsymbol{W}^2)-(\boldsymbol{W}\odot\boldsymbol{W}^2)\boldsymbol{W}-\boldsymbol{W}\odot\boldsymbol{W}^3-2\boldsymbol{W}^{\odot 2}\boldsymbol{W}^3-\mathrm{d}(\boldsymbol{W})^2\boldsymbol{W}\mathrm{d}(\boldsymbol{W}^2)\\
&-\boldsymbol{W}\odot\boldsymbol{W}^2\odot\boldsymbol{W}^2-\boldsymbol{W}\boldsymbol{W}^{\odot 3}\boldsymbol{W}-2\boldsymbol{W}\odot\boldsymbol{W}^2\odot\boldsymbol{W}^2-4\boldsymbol{W}^{\odot 5}\\
g_c(\boldsymbol{W},\boldsymbol{R})=&\boldsymbol{P}\mathrm{d}(\boldsymbol{P}^4)+\mathrm{d}(\boldsymbol{P}^4)\boldsymbol{P}+\boldsymbol{P}^2\mathrm{d}(\boldsymbol{P}^3)+\mathrm{d}(\boldsymbol{P}^3)\boldsymbol{P}^2+\boldsymbol{P}\mathrm{d}(\boldsymbol{P}^2)\boldsymbol{P}^2+\boldsymbol{P}^2\mathrm{d}(\boldsymbol{P}^2)\boldsymbol{P}+\boldsymbol{P}\mathrm{d}(\boldsymbol{P}^3)\boldsymbol{P}\\
&+\boldsymbol{P}^2\odot\boldsymbol{P}^{\odot 3}+3\boldsymbol{P}\odot(\boldsymbol{P}^{\odot 2})^2+2\boldsymbol{P}\mathrm{d}(\boldsymbol{P}^2)\odot\boldsymbol{P}^{\odot 2}+2\mathrm{d}(\boldsymbol{P}^2)\boldsymbol{P}\odot\boldsymbol{P}^{\odot 2}\\
&+4\mathrm{d}(\boldsymbol{P}^2)\boldsymbol{P}^{\odot 3}+4\boldsymbol{P}^{\odot 3}\mathrm{d}(\boldsymbol{P}^2)-\boldsymbol{P}\mathrm{d}(\boldsymbol{P}\mathrm{d}(\boldsymbol{P}^2)\boldsymbol{P})-\mathrm{d}(\boldsymbol{P}\mathrm{d}(\boldsymbol{P}^2)\boldsymbol{P})\boldsymbol{P}\\
&-2\boldsymbol{P}(\boldsymbol{P}^{\odot 2}\odot\boldsymbol{P}^2)-2(\boldsymbol{P}^{\odot 2}\odot\boldsymbol{P}^2)\boldsymbol{P}-\boldsymbol{P}^{\odot 2}\boldsymbol{P}^2-\boldsymbol{P}^2\boldsymbol{P}^{\odot 2}\\
&-2\boldsymbol{P}\mathrm{d}(\boldsymbol{P}^2)^2-2\mathrm{d}(\boldsymbol{P}^2)^2\boldsymbol{P}-\boldsymbol{P}(\boldsymbol{P}\odot\boldsymbol{P}^2)-(\boldsymbol{P}\odot\boldsymbol{P}^2)\boldsymbol{P}-\boldsymbol{P}\odot\boldsymbol{P}^3-2\boldsymbol{P}^{\odot 2}\boldsymbol{P}^3-\mathrm{d}(\boldsymbol{P})^2\boldsymbol{P}\mathrm{d}(\boldsymbol{P}^2)\\
&-\boldsymbol{P}\odot\boldsymbol{P}^2\odot\boldsymbol{P}^2-\boldsymbol{P}\boldsymbol{P}^{\odot 3}\boldsymbol{P}-2\boldsymbol{P}\odot\boldsymbol{P}^2\odot\boldsymbol{P}^2-4\boldsymbol{P}^{\odot 5}
\end{aligned}$$

The computational time complexity of the previous cases for f_c and g_c are $O(r(n))$ and $O(r(dn))$, respectively, since computing f_c by the formula above only requires standard matrix operations between $n\times n$ matrices, and computing g_c by the formula above only requires standard matrix operations between $dn\times dn$ matrices. For the case $c\geq 7$, [47] gives an estimation on the upper bound of the computational time complexity as $O(n^{\lceil (c+3)/2\rceil})$.

C. Main Theory

We formulate theory for adversarial corruption in Section C.1 and for the uniform corruption model in Section C.2. The latter theory extends the one stated in Section 4.

Both settings use the following common notation. Let E_g be the set of good (clean) edges, E_b be the set of bad (corrupted) edges, and N_{ij}^c be the set of simple c-cycles containing ij. Let G_{ij}^c be the set of good simple c-cycles with respect to ij. That is, for any cycle $L\in G_{ij}^c$, L is simple of length c and $L\setminus\{ij\}$ are all clean.

C.1. Theory for Adversarial Corruption

In this section we focus on the adversarial corruption model [25]. The adversarial corruption model makes no assumption on the graph topology or the corrution pattern. The only assumption is that for each $ij\in E_g$, $g_{ij}=g_{ij}^*$, and for each $ij\in E_b$, $g_{ij}\neq g_{ij}^*$. Since LongSync is a modified and vectorized version of CEMP for higher-order cycles, it inherits the robustness of CEMP to adversarial corruption. Define $\lambda=\max_{ij\in E}|B_{ij}^c|/|N_{ij}^c|$ where $B_{ij}^c=N_{ij}^c\setminus G_{ij}^c$ is the set of bad cycles with respect to ij (namely at least one of the other $(c-1)$ edges in the cycle are corrupted). In the scenario of adversarial corruption with an assumption on λ, we can guarantee linear convergence of LongSync as follows.

Theorem C.1. *Assume data is generated by the adversarial corruption model with $\lambda<\frac{1}{1+(c-1)^2}$. Assume the parameters $\{\beta_t\}_{t=1}^{t_{\max}}$ of LongSync with c-cycles satisfy $\beta_0\leq 1/(c-1)$, $\beta_{t+1}=r\beta_t$ and $1<r<\frac{1}{c-1}\sqrt{\frac{1-\lambda}{\lambda}}$. Then the corruption levels $\{s_{ij}^{(t)}\}_{ij\in E}$ estimated by LongSync satisfy the following equation:*

$$\max_{ij\in E}|s_{ij}^{(t)}-s_{ij}^*|\leq\frac{1}{(c-1)\beta_0 r^t}\text{ for all } t\geq 0. \tag{21}$$

Proof. Let $\epsilon_{ij}(t)=|s_{ij}^{(t)}-s_{ij}^*|$ and $\epsilon(t)=\max_{ij\in E}\epsilon_{ij}(t)$. By the fact that $|d_L-s_{ij}^*|\leq s_L^*$, $G_{ij}^c\subseteq N_{ij}^c$ and $s_L^*=0$ for $L\in G_{ij}^c$, we obtain that

$$\begin{aligned}
(\epsilon_{ij}(t+1))^2=|s_{ij}^{(t)}-s_{ij}^*|^2&=\left|\sqrt{\frac{\sum_{L\in N_{ij}^c}e^{-\beta_t s_L^{(t)}}d_L^2}{\sum_{L\in N_{ij}^c}e^{-\beta_t s_L^{(t)}}}}-s_{ij}^*\right|^2\\
&\leq\frac{\sum_{L\in N_{ij}^c}e^{-\beta_t s_L^{(t)}}|d_L-s_{ij}^*|^2}{\sum_{L\in N_{ij}^c}e^{-\beta_t s_L^{(t)}}}\\
&\leq\frac{\sum_{L\in N_{ij}^c}e^{-\beta_t s_L^{(t)}}(s_L^*)^2}{\sum_{L\in N_{ij}^c}e^{-\beta_t s_L^{(t)}}}\\
&\leq\frac{\sum_{L\in B_{ij}^c}e^{-\beta_t s_L^{(t)}}(s_L^*)^2}{\sum_{L\in G_{ij}^c}e^{-\beta_t s_L^{(t)}}}\\
&\leq\frac{\sum_{L\in B_{ij}^c}e^{-\beta_t\sum_{e\in L}\epsilon_e(t)}(s_L^*)^2}{\sum_{L\in G_{ij}^c}e^{-\beta_t\sum_{e\in L}\epsilon_e(t)}}\\
&\leq\frac{1}{|G_{ij}^c|}e^{2\beta_t(c-1)\epsilon(t)}\sum_{L\in B_{ij}^c}e^{-\beta_t s_L^*}(s_L^*)^2.
\end{aligned} \tag{22}$$

We prove the theorem by induction. Note that the case $t=0$ is equivalent to $\epsilon(0)\leq 1/(c-1)\beta_0$, and this immediately follows from the fact that $0\leq\epsilon_{ij}(0)\leq 1$ and the assumption $\beta_0<1/(c-1)$. We next prove $\epsilon(t+1)<1/(c-1)\beta_{t+1}$ from $\epsilon(t)<1/(c-1)\beta_t$. By the inequality above, the induction assumption, the fact that $x^2e^x<4/(ax)^2$ with $x=s_L^*$ and $a=\beta_t$ and the definition of λ and r we have

$$(\epsilon_{ij}(t+1))^2\leq\frac{1}{|G_{ij}^c|}\cdot e^2\cdot\frac{4|B_{ij}^c|}{e^2\beta_t^2}=\frac{4|B_{ij}^c|}{|G_{ij}^c|\beta_t^2}\leq\frac{4\lambda}{(1-\lambda)\beta_t^2}=\frac{1}{\beta_t^2r^2(c-1)^2}=\frac{1}{\beta_{t+1}^2(c-1)^2}. \tag{23}$$

The theorem follows by taking the maximum of the left hand side and then the square root of both sides of the above equation. □

C.2. Theory for Uniform Corruption Model

Throughout the rest of the paper we use $P(A)$ to denote the probability of event A. Let $p_0=P(g_{ij}=g_{ij}^*)$ for each edge $ij\in E_b$. By the choice of corruption model, p_0 only depends on the group $\mathcal{G}$. Let $q_*=1-q+qp_0=P(ij\in E_g|ij\in E)$. Let $q_g=1-q$. We remark that for rotation synchronization (in fact any Lie group synchronization), $q_g=q_*$ and $p_0=0$.

Recall for each $e\in E$, s_e^* is the ground truth corruption level of edge e. For $L=(ik_1,k_1k_2,\cdots,k_{c-2}j)\in N_{ij}^c$, we denote $s_L^*=\sum_{e\in L\setminus\{ij\}}s_e^*$. To state our main theorem, we let $\mathcal{F}(\beta)=\{f_\tau(x):=e^{-\tau x+2}\tau^2x^2/4:\tau>\beta\}$ and $V(\beta)=\sup_{\tau>\beta}\mathrm{Var}(f_\tau(s_L^*))$. Due to the model assumptions, the distribution of $f_\tau(s_L^*)$ is independent of the choice of $L\in N_{ij}^c$.

Using the above notation, we formulate the following theorem, which generalizes Theorem 4.1

Theorem C.2. *Let $0<r<1$, $0<q<1$, $0<p\leq1$. Assume we use LongSync with cycles of length c and $n/\log n=\Omega((pq_g)^{-\frac{c-1}{c-2-\epsilon}})$ for some $\epsilon>0$. Assume*

$$0<\frac{1}{\beta_0}<\frac{q_g^{c-1}q_*^{c-1}}{16(1-q_*^{c-1})(c-1)^2\beta_1}, \tag{24}$$

$$V(\beta_1)<\frac{r}{16(c-1)}\cdot\frac{q_*^{c-1}}{1-q_*^{c-1}}, \tag{25}$$

$$1/\beta_{t+1}=r/\beta_t \text{ for all } t\geq1, \tag{26}$$

$$\min(np,n^{c-2-\epsilon}p^{c-1})\gtrsim\frac{(1-q_*^{c-1})^2}{q_*^{2(c-1)}r^2}. \tag{27}$$

Then with probability at least $1-4cn^2\exp\left(-K\eta_0^2(pq_*)^{\frac{c-1}{c-2}}n\right)-2e^2c\cdot\exp\left(-n^{\epsilon/(c-1)}+c\log n\right)-n^2\exp\left(-\frac{\ln2}{2}\min(np,n^{c-2-\epsilon}p^{c-1})V(\beta_1)\right)-2n^2\cdot\exp\left(-\frac{\eta e_{\mathcal{G}}}{8c}\ln(1+\frac{e_{\mathcal{G}}}{2(c-1)\beta_0v_{\mathcal{G}}})\min(np,n^{c-2-\epsilon}p^{c-1})\right)$, *where $\eta_0,\eta,K,e_{\mathcal{G}},v_{\mathcal{G}}$ are absolute constants, we have* $\max_{ij\in E}|s_{ij}^*-s_{ij}^{(t)}|\leq\frac{1}{2c\beta_t}$ *for all* $t\geq1$.

Remark C.3. As is shown in [25], for $\mathcal{G}\in SO(3)$, $V(\beta)\sim O(\beta^{-3})$. Therefore $n/\log n\sim p^{-(c-1)/(c-2-\epsilon)}q_g^{-7(c-1)/3(c-2-\epsilon)}$ is the minimal sample complexity dependence for $\mathcal{G}=SO(3)$ such that with high probability, the conclusion of Theorem 4.1 holds true.

C.3. Proof of Theorem C.2

We adopt the proof framework of [25]. The major difficulty of the proof is the dependence in the cycle inconsistency measures of cycles in N_{ij}^c when $c\geq4$. For example, the cycle inconsistency measure of a 4-cycle $L_1=(ik_1,k_1k_2,k_2j)$ is not independent with that of $L_2=(ik_1,k_1k_3,k_3j)$, while for a pair of 3-cycles their ratios are always independent. This means that the required concentration inequalities cannot be obtained by directly applying the standard Chernoff bounds. Nonetheless, we have integrated various mathematical techniques from [3, 7, 22–24, 48] to derive Theorem 4.1, which offers improvements over theorem 7 presented in [25].

For convenience for any $c\geq3$, we define a c-path as a path that involves c vertices, and we define an ij,c-path as a c-path that starts from i and ends at j. We extend the definition of N_{ij}^c as the set of ij,c-paths in graph G.

We first prove that with high probability, the number of c_1-cycles concentrates around its mean for any $c_1\leq c$. More specifically, let $n_{c_1}=(n-2)(n-3)(n-4)\cdots(n-c_1+1)$ be the number of possible ij,c_1-path candidates, and $m_{c_1}=\max(p^{c_1-1}n_{c_1},n^\epsilon)$. Therefore the expected number of ij,c_1-paths is $p^{c_1-1}n_{c_1}$. For any $\epsilon,\eta>0$ we define the (ϵ,η_0)-regular Erdős-Rényi graph condition as follows:

Definition C.4. Let $\delta=\sup\{\delta>0 \text{ s.t. } np^{1+\delta}/\log n\to\infty\}$ and $c_0=\lceil2+\delta^{-1}\rceil$. A graph G satisfies the (ϵ,η_0)-regular Erdős-Rényi graph condition if and only if the following conditions hold true:

- For any $i\neq j\in[n]$ and $c_1\geq c_0$,

$$(1-\eta_0)m_{c_1}<|N_{ij}^{c_1}|<(1+\eta_0)m_{c_1} \tag{28}$$

  and

$$(1-\eta_0)q_*^{c_1-1}m_{c_1}<|G_{ij}^{c_1}|<(1+\eta_0)q_*^{c_1-1}m_{c_1}; \tag{29}$$

- For any $i \neq j \in [n]$ and $c_1 < c_0$,

$$0 \leq |N_{ij}^{c_1}| < m_{c_1}. \tag{30}$$

.

We have the following theorem on the phase transition of the number of c-paths:

Theorem C.5. *Assume G is generated with the uniform corruption model UCM(n,p,q), and $\epsilon,\eta > 0$ are constants. Then the (ϵ,η_0)-regular E-R graph condition holds with probability at least $1 - cn^2\exp(-\frac{\eta_0^2}{5c}pn) - cn^2\exp(-K\eta_0^2 p^{\frac{c-1}{c-2}}n) - cn^2\exp(-\frac{\eta_0^2}{5c}pq_* n) - cn^2\exp\left(-K\eta_0^2(pq_*)^{\frac{c-1}{c-2}}n\right) - 2e^2cn^2\exp\left(-n^{\epsilon/(c-1)} + (c-2)\log n\right)$, which is almost 1 by the condition $n/\log n = \Omega((pq_g)^{-\frac{c-1}{c-2-\epsilon}})$.*

The proof of Theorem C.5 is put in section D. Based on this theorem, we have a concentrated 'initialization' of corruption level estimates after the first iteration:

Theorem C.6. *(Initialization) Assume the (ϵ,η_0)-regular E-R graph condition holds. Recall that the corruption level estimation of LongSync with cycle length c at $t=0$ is*

$$s_{ij}^{(0)} = \sqrt{\frac{\sum_{L \in N_{ij}^c} d_L^2}{|N_{ij}^c|}}. \tag{31}$$

Denote $e_{\mathcal{G}} = \mathbb{E}d_L^2$ and $v_{\mathcal{G}} = Var(d_L^2)$. Then for any $\eta > 0$ and $ij \in E$,

$$P(|(s_{ij}^{(0)})^2 - \mathbb{E}(s_{ij}^{(0)})^2| > \eta\mathbb{E}(s_{ij}^{(0)})^2) < 2\exp\left(-\frac{\eta e_{\mathcal{G}}}{8c}\ln(1+\frac{\eta e_{\mathcal{G}}}{2v_{\mathcal{G}}})\min(np, n^{c-2-\epsilon}p^{c-1})\right). \tag{32}$$

Let $\lambda = \max_{ij\in E}|B_{ij}^c|/|N_{ij}^c|$ where $B_{ij}^c = N_{ij}^c \setminus G_{ij}^c$ is the set of bad ij,c-paths. To prove the linear convergence, we need the following three lemmas:

Lemma C.7. If $\max_{ij\in E}|(s_{ij}^{(0)})^2 - \mathbb{E}(s_{ij}^{(0)})^2| \leq \frac{1}{2(c-1)\beta_0}$, then

$$\max_{ij\in E}|s_{ij}^{(1)} - s_{ij}^*| \leq \frac{\lambda}{1-\lambda}\frac{2(c-1)}{q_g^{c-1}\beta_0}. \tag{33}$$

Lemma C.8. Assume that $\max_{ij\in E}|s_{ij}^{(1)} - s_{ij}^*| < 1/(2(c-1)\beta_1)$, $\beta_t = r\beta_{t+1}$ for $t \geq 1$, and

$$\max_{ij\in E}\frac{1}{|B_{ij}^c|}\sum_{L\in B_{ij,c}} e^{-\beta_t s_L^*}(s_L^*)^2 < \frac{1}{M\beta_t^2} \text{ for all } t \geq 1, \tag{34}$$

where $M = 4(c-1)^2 e\lambda/((1-\lambda)r^2)$. Then the LongSync corruption level estimates satisfy

$$\max_{ij\in E}|s_{ij}^{(t)} - s_{ij}^*| < \frac{1}{\beta_1}r^{t-1} \text{ for all } t \geq 1. \tag{35}$$

Lemma C.9. If either s_{ij}^* for $ij \in E_b$ is supported on $[a,\infty)$ and $a \geq 1/|B_{ij}^c|$ or Q is differentiable and $Q'(x)/Q(x) \lesssim 1/x$ for $x < P(1)$, then there exists an absolute constant K'' such that

$$\begin{aligned} P\Bigg(&\sup_{f_\tau\in\mathcal{F}(\beta)}\frac{1}{|B_{ij}^c|}\sum_{L\in B_{ij}^c} f_\tau(s_L^*) > V(\beta) \\ &+ K''\sqrt{\frac{\log\min(np, n^{c-2-\epsilon}p^{c-1})}{\min(np, n^{c-2-\epsilon}p^{c-1})}}\Bigg) \\ < &\exp\left(-\frac{\ln 2}{2}\min(np, n^{c-2-\epsilon}p^{c-1})V(\beta)\right). \end{aligned} \tag{36}$$

where $\mathcal{F}(\beta) = \{f_\tau(x) = e^{-\tau x + 2}\tau^2 x^2/4 : \tau > \beta\}$.

Lemma C.7 and C.8 are direct extensions of lemma 4 and lemma 5 of [25]. Lemma C.9, however, involves the extension of theorem 2.3 in [3] to the supremum of locally independent empirical processes and Hajnal-Szemerédi theorem for equitable coloring. We refer the reader to section D for the proof of these lemmas.

Proof of the main theorem. By the regular E-R graph condition, we can choose appropriate η_0 so that

$$\frac{1}{4}\frac{q_*^{c-1}}{1-q_*^{c-1}} < \frac{1-\lambda}{\lambda} < 4\frac{q_*^{c-1}}{1-q_*^{c-1}}. \tag{37}$$

To guarantee the condition (34) of Lemma C.8, we need to choose β_1 such that $V(\beta_1) < e/2M$ and n large enough such that $\log(\min(np,\ n^{c-2-\epsilon}p^{c-1}))/\ \min(np,\ n^{c-2-\epsilon}p^{c-1}) < e^2/4K''^2M^2$. By the assumption that $V(\beta_1) < (rq_*^{c-1})/16(c-1)(1-q_*^{c-1})$, $M = 4(c-1)^2 e\lambda/((1-\lambda)r^2)$ and (37) we know that $V(\beta_1) < e/2M$. By the assumption that $\min(np, n^{c-2-\epsilon}p^{c-1}) \gtrsim (1-q_*^{c-1})^2/q_*^{2(c-1)}r^2$ we know that $\log(\min(np, n^{c-2-\epsilon}p^{c-1}))/\min(np, n^{c-2-\epsilon}p^{c-1}) < e^2/4K''^2M^2$. Therefore the condition (34) of Lemma C.8 holds true.

On the other hand, by Theorem C.6 with $\eta = 1/2(c-1)\beta_0$ we know that w.h.p. the condition of Lemma C.7 holds true. By the assumption that $1/\beta_0 < q_*^{c-1}q_g^{c-1}/16(1-q_*^{c-1})(c-1)^2\beta_1$, we know that the conclusion of Lemma C.7 implies the first assumption of Lemma C.8.

Therefore, the proof of the theorem follows from the conclusion of Lemma C.8. □

D. Proofs of Auxiliary Results

We provide additional results for auxiliary theorems and lemmata used in the previous section.

Proof of Theorem C.5. We have the following basic lemmas:

Lemma D.1. (Concentration of number of paths of length $\geq c_0 - 1$ with fixed endpoints) Let $0 \leq q < 1$, $0 < p \leq 1$, $n \in \mathbb{N}$ with $np \geq \Theta(1)$. Assume data is generated by UCM(n,p,q), and $c \geq c_0$. For any $\eta_0 > 0$, there exists a constant $K > 0$ that only depends on c, such that

$$P(|N_{ij}^c| - p^{c-1}n_c < \eta_0 p^{c-1}n_c) < \exp(-\frac{\eta_0^2}{5c}pn) \tag{38}$$

$$P(|N_{ij}^c| - p^{c-1}n_c > \eta_0 p^{c-1}n_c) < \exp(-K\eta_0^2 p^{\frac{c-1}{c-2}} n) \tag{39}$$

for any fixed $i \neq j \in V$, and

$$P(|N_{ij}^c| - p^{c-1}n_c < \eta_0 p^{c-1}n_c) < |E|\exp(-\frac{\eta_0^2}{5c}pn) \tag{40}$$

$$P(|N_{ij}^c| - p^{c-1}n_c > \eta_0 p^{c-1}n_c) < |E|\exp(-K\eta_0^2 p^{\frac{c-1}{c-2}} n). \tag{41}$$

Proof. Let $M_{ij}^c = \{(i,k_1,k_2,\cdots,k_{c-2},j) : i,k_1,k_2,\cdots,k_{c-2},j \in [n] \text{ are different}\}$. Note that $|N_{ij}^c| = \sum_{\alpha\in M_{ij}^c} I_\alpha$, where $I_\alpha = 1_{ik_1\in E}1_{k_1k_2\in E}\cdots 1_{k_{c-3}k_{c-2}\in E}1_{k_{c-2}j\in E}$ for $\alpha = (i,k_1,k_2,\cdots,k_{c-2},j)$. For any $\alpha,\beta \in M_{ij}^c$, define $\omega = \sum_{\alpha\in M_{ij}^c}\mathbb{E}I_\alpha = \sum_{\alpha\in M_{ij}^c}p^{c-1} = p^{c-1}n_c$. Let us write $\alpha\sim\beta$ if α, $\beta \in M_{ij}^c$ with at least one common edge, and define $\delta = (\sum_{\alpha\sim\beta}\mathbb{E}I_\alpha I_\beta)/\omega$. (This sum should be interpreted as the sum over all pairs (α,β), so each pair is counted twice.) By theorem 1 of [22], we have the following inequality:

$$P(|N_{ij}^c| < (1-\eta_0)p^{c-1}n_c) \leq \exp(-\frac{\eta_0^2\omega}{2(1+\delta)}). \tag{42}$$

Denote $|\alpha\backslash\beta|$ as the number of nodes that belong to β but do not belong to α. By the definition of δ, we have the following estimate:

$$
\begin{aligned}
\delta &= (\sum_{\alpha\sim\beta}\mathbb{E}I_\alpha I_\beta)/\omega \\
&= \frac{1}{\omega}\sum_{\alpha\in M_{ij}^c}\sum_{k=1}^{c-3}\sum_{\alpha\sim\beta \text{ and } |\alpha\backslash\beta|=k}\mathbb{E}I_\alpha I_\beta \\
&= \frac{|M_{ij}^c|}{\omega}\sum_{k=1}^{c-3}\sum_{\alpha\sim\beta \text{ and } |\alpha\backslash\beta|=k} p^{k+c-1} \\
&\leq \frac{(n-2)(n-3)\cdots(n-c+1)}{p^{c-1}(n-2)(n-3)\cdots(n-c+1)}\sum_{k=1}^{c-3}(n-2)(n-3)\cdots(n-k-1)p^{k+c} \\
&\leq \frac{1}{p^{c-1}}c(n-2)(n-3)\cdots(n-c+2)p^{2c-3} \\
&\leq c(n-2)(n-3)\cdots(n-c+2)p^{c-2} = \frac{c\omega}{(n-c+1)p}.
\end{aligned} \tag{43}
$$

Plugging (43) to (42) gives:

$$
\begin{aligned}
P(|N_{ij}^c| < (1-\eta_0)p^{c-1}n_c) &\leq \exp(-\frac{\eta_0^2\omega}{2(1+\delta)}) \\
&< \exp(-\frac{\eta_0^2\omega}{4\delta}) \\
&\leq \exp(-\frac{\eta_0^2\omega(n-c+1)p}{4c\omega}) \\
&< \exp(-\frac{\eta_0^2 np}{5c}).
\end{aligned} \tag{44}
$$

Therefore inequality (38) is proved, and inequality (40) follows from a union bound argument.

For the upper tail, let A be an arbitrary subset of $\{k_1,k_2,\cdots,k_{c-2}\}$, the set of free vertices of an ij,c-path. Denote $\mathbb{M}_A$ as the expected number of ij,c-paths $(ik_1,k_1k_2,\cdots,k_{c-2}j)$, where the vertices in A are fixed, and let $\mathbb{M}_k = \max_{|A|\geq k}\mathbb{M}_A$. We have the following calculation:

$$
\mathbb{M}_k = \begin{cases} n^{c-2-k}p^{c-1-k}, & k\leq c-3 \\ 1, & k=c-2 \end{cases}. \tag{45}
$$

Let $\lambda = \eta_0^2(n-c+1)p^{\frac{c-1}{c-2}}$. By $c\geq c_0$, we know that $\lambda=\omega(\log n)$. Also, by setting $M_0=\mathbb{M}_0$ and $M_k = M_0\lambda^{-k}$ we know that for all $0\leq k\leq c-2$, $M_k\geq\mathbb{M}_k$. Therefore we can apply theorem 1.2 in [48] and get the following inequality

$$
P(|N_{ij}^c| - p^{c-1}n_c > \eta_0 n_c) \leq \exp(-K_0\eta_0^2(n-c+1)p^{\frac{c-1}{c-2}}) \tag{46}
$$

where K_0 is a constant that only depends on c. Let $K=K_0/2$. By the order of c we know that

$$
P(|N_{ij}^c| - p^{c-1}n_c > \eta_0 n_c) \leq \exp(-K\eta_0^2 np^{\frac{c-1}{c-2}}). \tag{47}
$$

Therefore inequality (39) is proved, and inequality (41) follows from a union bound argument. □

Lemma D.2. Let $0\leq q<1$, $0<p\leq 1$, $n\in\mathbb{N}$ with $np\geq\Theta(1)$. Assume data is generated by UCM(n,p,q), $c\geq c_0$, and K is the constant in Lemma D.1. For any $\eta_0>0$, we have

$$
P(|G_{ij}^c| - p^{c-1}q_*^{c-1}n_c < \eta_0 p^{c-1}q_*^{c-1}n_c) < \exp(-\frac{\eta_0^2}{5c}pq_*n) \tag{48}
$$

$$
P(|G_{ij}^c| - p^{c-1}q_*^{c-1}n_c > \eta_0 p^{c-1}q_*^{c-1}n_c) < \exp(-K\eta_0^2 pq_*n) \tag{49}
$$

for any fixed $i \neq j \in V$, and

$$P(|G_{ij}^c| - p^{c-1}q_*^{c-1}n_c < \eta_0 p^{c-1}q_*^{c-1}n_c) < |E|\exp(-\frac{\eta_0^2}{5c}pq_*n) \tag{50}$$

$$P(|G_{ij}^c| - p^{c-1}q_*^{c-1}n_c > \eta_0 p^{c-1}q_*^{c-1}n_c) < |E|\exp(-K\eta_0^2 pq_*n). \tag{51}$$

Lemma D.2 is proved by replacing p with pq_* in the proof of Lemma D.1.

To count the shorter paths which has a vanishing expectation when n tends to infinity, we need the following concentration inequality:

Lemma D.3. (Concentration of number of paths with length $\leq c_0 - 2$) Let $0 \leq q < 1$, $0 < p \leq 1$, $n \in \mathbb{N}$ with $np \geq \Theta(1)$. Assume data is generated by UCM(n,p,q), and $c < c_0$. For any $\epsilon > 0$, there exists a constant $K' > 0$ that only depends on c, such that

$$P(|N_{ij}^c| > K'n^\epsilon) < 2e^2\exp(-n^{\epsilon/(c-1)} + (c-2)\log n) \tag{52}$$

for any fixed $i \neq j \in V$, and

$$P(|N_{ij}^c| > K'n^\epsilon) < 2e^2|E|\exp(-n^{\epsilon/(c-1)} + (c-2)\log n). \tag{53}$$

Proof. Define the multivariable polynomial $f(\{x_{pq}\}_{p\neq q\in[n]}) = \sum_{\alpha\in M_{ij}^c} x_\alpha$, where $x_\alpha = x_{ik_1}x_{k_1k_2}\cdots x_{k_{c-2}j}$ for $\alpha = (i,k_1,k_2,\cdots,k_{c-2},j)$ in $M_{ij}^c = \{(i,k_1,k_2,\cdots,k_{c-2},j) : i,k_1,k_2,\cdots,k_{c-2},j \in [n]$ are different$\}$. Note that $|N_{ij}^c| = f(\{1_{pq\in E}\}_{p\neq q\in[n]})$. Let $A \subseteq \{x_{pq\in E} : p \neq q \in [n]\}$ be a subset of the variables of f, and $f_A(\{x_{pq}\}_{p\neq q\in[n]})$ be the partial derivative of $f(\{x_{pq}\}_{p\neq q\in[n]})$ with respect to all variables in A. Let $\partial_A|N_{ij}^c| = f_A(\{1_{pq\in E}\}_{p\neq q\in[n]})$. Define $E_k = \max_{|A|\geq k}\mathbb{E}(\partial_A|N_{ij}^c|)$. By the main theorem in [24], we know that

$$P(|N_{ij}^c - E_0| > K'n^{(c-1)\epsilon}\sqrt{E_0E_1}) < 2e^2\exp(-n^\epsilon + (c-2)\log n). \tag{54}$$

Because $c < c_0$, we know that for any $k \in \mathbb{N}$, $\max_{|A|\leq c-2}\mathbb{E}(\partial_A|N_{ij}|) = o(1)$ and $\max_{|A|=c-1}\mathbb{E}(\partial_A|N_{ij}|) = 1$. Therefore, $E_0 = E_1 = 1$. Plugging these values into inequality (54) and substituting ϵ with $\epsilon/(c-1)$ results in inequality (52). Inequality (53) is obtained from a union probability bound argument. □

With the estimates above, the regular E-R graph condition holds with probability at least $1 - n^2\exp(-\frac{\eta_0^2}{5c}pn) - n^2\exp(-K\eta_0^2 p^{\frac{c-1}{c-2}}n) - n^2\exp(-\frac{\eta_0^2}{5c}pq_*n) - n^2\exp(-K\eta_0^2(pq_*)^{\frac{c-1}{c-2}}n) - 2e^2n^2\exp(-n^\epsilon + (c-2)\log n)$.

□

Proof of Theorem C.6. For any $L \in N_{ij}^c$ and $pq \in L$, we say L' is correlated with L if $L \cap L'$ is nonempty, and L' is correlated with $L\backslash\{pq\}$ if $(L\backslash\{pq\}) \cap L'$ is nonempty. We denote C_L as the set of ij,c-paths in N_{ij}^c that is correlated with L, and denote $C_{L\backslash\{pq\}}$ as the set of ij,c-paths in N_{ij}^c that is correlated with $L\backslash\{pq\}$. With the regular E-R graph condition, we know that for any $L \in N_{ij}^c$,

$$|C_L| \leq \sum_{pq\in L}|C_{L\backslash\{pq\}}| \tag{55}$$

$$\leq m_{c-1} + m_1m_{c-2} + m_2m_{c-3} + \cdots + m_{c-2}m_1 + m_{c-1} \tag{56}$$

$$< cm_{c-1}. \tag{57}$$

Denote $\Delta_1 = \max_{L\in N_{ij}^c}|C_L|$. Then we know that $\Delta_1 < cm_{c-1} < c\max(n^\epsilon, n^{c-3}p^{c-2})$. We apply theorem 2.5 in [23] on $\sum_{L\in N_{ij}^c} d_L^2$ and $\sum_{L\in N_{ij}^c}(-d_L^2)$ and get the following inequalities:

$$P(\sum_{L\in N_{ij}^c} d_L^2 > (1+\eta)\mathbb{E}\sum_{L\in N_{ij}^c} d_L^2) < \exp(-\frac{|N_{ij}^c|v_\mathcal{G}}{\Delta_1}\varphi(\frac{\eta\mathbb{E}\sum_{L\in N_{ij}^c} d_L^2}{|N_{ij}^c|v_\mathcal{G}(1+\Delta_1/8|N_{ij}^c|)})) \tag{58}$$

and

$$P(\sum_{L\in N_{ij}^c} d_L^2 < (1-\eta)\mathbb{E}\sum_{L\in N_{ij}^c} d_L^2) < \exp(-\frac{|N_{ij}^c|v_\mathcal{G}}{\Delta_1}\varphi(\frac{\eta\mathbb{E}\sum_{L\in N_{ij}^c} d_L^2}{|N_{ij}^c|v_\mathcal{G}(1+\Delta_1/8|N_{ij}^c|)})) \tag{59}$$

where $\varphi(x) = (1+x)\ln(1+x) - x$. Note that $\varphi(x) \geq x\ln(1+x)/2$ for any $x \geq 0$. By the regular E-R graph condition we have $|N_{ij}^c| \geq (1-\eta_0)n^{c-2}p^{c-1}$, and therefore $\Delta_1/|N_{ij}^c| \leq \max(1/(n^{c-2}p^{c-1}), 1/(np))/(1-\eta_0) < 1$. Also, since all the d_L^2's for $L \in N_{ij}^c$

follow the same distribution with mean $e_{\mathcal{G}}$ and variance $v_{\mathcal{G}}$, we know that $\mathbb{E}\sum_{L\in N_{ij}^c} d_L^2 = |N_{ij}^c| e_{\mathcal{G}}$. Therefore RHS of (58) and (59) can be upper bounded as follows:

$$\begin{aligned}
\text{RHS of (58) and (59)} \leq & \exp\Bigg(-\frac{|N_{ij}^c|v_{\mathcal{G}}}{\Delta_1}\cdot\frac{\eta\mathbb{E}\sum_{L\in N_{ij}^c} d_L^2}{2|N_{ij}^c|v_{\mathcal{G}}(1+\Delta_1/8|N_{ij}^c|)} \\
&\cdot\ln(1+\frac{\eta\mathbb{E}\sum_{L\in N_{ij}^c} d_L^2}{|N_{ij}^c|v_{\mathcal{G}}(1+\Delta_1/8|N_{ij}^c|)})\Bigg) \\
=& \exp\Bigg(-\frac{1}{\Delta_1}\cdot\frac{\eta|N_{ij}^c|e_{\mathcal{G}}}{2(1+\Delta_1/8|N_{ij}^c|)}\ln(1+\frac{\eta e_{\mathcal{G}}}{v_{\mathcal{G}}(1+\Delta_1/8|N_{ij}^c|)})\Bigg) \\
\leq & \exp\Bigg(-\frac{\eta e_{\mathcal{G}}|N_{ij}^c|}{4\Delta_1}\ln(1+\frac{\eta e_{\mathcal{G}}}{2v_{\mathcal{G}}})\Bigg) \\
\leq & \exp\Bigg(-\frac{\eta e_{\mathcal{G}}(1-\eta_0)n^{c-2}p^{c-1}}{4\max(n^{\epsilon},n^{c-3}p^{c-2})}\ln(1+\frac{\eta e_{\mathcal{G}}}{2v_{\mathcal{G}}})\Bigg) \\
\leq & \exp\Bigg(-\frac{\eta e_{\mathcal{G}}}{8c}\ln(1+\frac{\eta e_{\mathcal{G}}}{2v_{\mathcal{G}}})\min(np,n^{c-2-\epsilon}p^{c-1})\Bigg).
\end{aligned} \tag{60}$$

Combining the upper and lower tail bound together yields

$$P(|\sum_{L\in N_{ij}^c} d_L^2 - \mathbb{E}\sum_{L\in N_{ij}^c} d_L^2| > \eta\mathbb{E}\sum_{L\in N_{ij}^c} d_L^2) < 2\exp(-\frac{\eta e_{\mathcal{G}}}{8c}\ln(1+\frac{\eta e_{\mathcal{G}}}{2v_{\mathcal{G}}})\min(np,n^{c-2-\epsilon}p^{c-1})). \tag{61}$$

Then Theorem C.6 follows by (31). □

Proof of Lemma C.7. Denote $\gamma_{ij} = (s_{ij}^{(0)})^2 - \mathbb{E}(s_{ij}^{(0)})^2$ for $ij \in E$ and $\gamma = \max_{ij\in E}|\gamma_{ij}|$, so that the condition of the lemma can be written more simply as $1/2(c-1)\beta_0 \geq \gamma$. By rewriting $\mathbb{E}(s_{ij}^{(0)})^2$ as $q_g^{c-1}(s_{ij}^*)^2 + (1-q_g^{c-1})z_{\mathcal{G}} + \gamma_{ij}$ and invoking lemma 1 in [25] and equations (6) (7), we have the following bound:

$$\begin{aligned}
|s_{ij}^{(1)} - s_{ij}^*|^2 \leq & \frac{\sum_{L\in N_{ij}^c} e^{-\beta_0\sqrt{\sum_{e\in L} q_g^{c-1}(s_e^*)^2 + (1-q_g^{c-1})z_{\mathcal{G}} + \gamma_e}}|d_L - s_{ij}^*|^2}{\sum_{L\in N_{ij}^c} e^{-\beta_0\sum_{e\in L}\sqrt{q_g^{c-1}(s_e^*)^2+(1-q_g^{c-1})z_{\mathcal{G}}+\gamma_e}}} \\
\leq & \frac{\sum_{L\in B_{ij}^c} e^{-\beta_0\sum_{e\in L}\sqrt{q_g^{c-1}(s_e^*)^2+(1-q_g^{c-1})z_{\mathcal{G}}+\gamma_e}}(s_L^*)^2}{\sum_{L\in G_{ij}^c} e^{-\beta_0\sum_{e\in L}\sqrt{q_g^{c-1}(s_e^*)^2+(1-q_g^{c-1})z_{\mathcal{G}}+\gamma_e}}}
\end{aligned} \tag{62}$$

By first applying the facts: $|\gamma_e| \leq \gamma$ and $s_e^* = 0$ for $e \in L$ where $L \in G_{ij}^c$, and at last the inequality $xe^{-ax} \leq 1/(ea)$ with $x = \sum_{e\in L}(s_e^*)^2$ and $a = \beta_0 q_g^{c-1}/2$, we obtain that

$$\begin{aligned}
|s_{ij}^{(1)} - s_{ij}^*|^2 \leq & \frac{\sum_{L\in B_{ij}^c} e^{-\beta_0\sum_{e\in L}\sqrt{q_g^{c-1}(s_e^*)^2+(1-q_g^{c-1})z_{\mathcal{G}}-\gamma}}(s_L^*)^2}{|G_{ij}^c|e^{-\beta_0(c-1)\sqrt{(1-q_g^{c-1})z_{\mathcal{G}}+\gamma}}} \\
=& \frac{\sum_{L\in B_{ij}^c} e^{-\beta_0\sum_{e\in L}(\sqrt{q_g^{c-1}(s_e^*)^2+(1-q_g^{c-1})z_{\mathcal{G}}-\gamma}-\sqrt{(1-q_g^{c-1})z_{\mathcal{G}}+\gamma})}(s_L^*)^2}{|G_{ij}^c|} \\
\leq & \frac{\sum_{L\in B_{ij}^c} e^{-\beta_0\sum_{e\in L}(q_g^{c-1}(s_e^*)^2-2\gamma)/2}(s_L^*)^2}{|G_{ij}^c|} \\
\leq & \frac{e^{2\beta_0(c-1)\gamma}\sum_{L\in B_{ij}^c} e^{-\beta_0 q_g^{c-1}\sum_{e\in L}(s_e^*)^2/2}(c-1)\sum_{e\in L}(s_e^*)^2}{|G_{ij}^c|} \\
\leq & \frac{2(c-1)|B_{ij}^c|}{|G_{ij}^c|\beta_0 q_g^{c-1}}.
\end{aligned} \tag{63}$$

The lemma is concluded by applying the union bound on $ij \in E$ and taking the square root on both sides of the above inequality. $\square$

Proof of Lemma C.8. Let $\epsilon_{ij}(t) = |s_{ij}^{(t)} - s_{ij}^*|$ and $\epsilon(t) = \max_{ij\in E} \epsilon_{ij}(t)$. We prove this lemma, or equivalently $\epsilon(t) < 1/2(c-1)\beta_t$ for all $t \geq 1$, by induction. We first note that $\epsilon(1) < 1/4\beta_t$ is an assumption of the lemma. Next we show that $\epsilon(t+1) < 1/2(c-1)\beta_{t+1}$ if $\epsilon(t) < 1/2(c-1)\beta_t$.

By the fact that $|d_L - s_{ij}^*| \leq s_L^*$, $G_{ij}^c \subseteq N_{ij}^c$ and $s_L^* = 0$ for $L \in G_{ij}^c$, we obtain that

$$\begin{aligned}
\epsilon_{ij}(t+1)^2 = |s_{ij}^{(t)} - s_{ij}^*|^2 = &|\sqrt{\frac{\sum_{L\in N_{ij}^c} e^{-\beta_t s_L^{(t)}} d_L^2}{\sum_{L\in N_{ij}^c} e^{-\beta_t s_L^{(t)}}}} - s_{ij}^*|^2 \\
&\leq \frac{\sum_{L\in N_{ij}^c} e^{-\beta_t s_L^{(t)}} |d_L - s_{ij}^*|^2}{\sum_{L\in N_{ij}^c} e^{-\beta_t s_L^{(t)}}} \\
&\leq \frac{\sum_{L\in N_{ij}^c} e^{-\beta_t s_L^{(t)}} (s_L^*)^2}{\sum_{L\in N_{ij}^c} e^{-\beta_t s_L^{(t)}}} \\
&\leq \frac{\sum_{L\in B_{ij}^c} e^{-\beta_t s_L^{(t)}} (s_L^*)^2}{\sum_{L\in G_{ij}^c} e^{-\beta_t s_L^{(t)}}} \\
&\leq \frac{\sum_{L\in B_{ij}^c} e^{-\beta_t \sum_{e\in L} \epsilon_e(t)} (s_L^*)^2}{\sum_{L\in G_{ij}^c} e^{-\beta_t \sum_{e\in L} \epsilon_e(t)}} \\
&\leq \frac{1}{|G_{ij}^c|} \sum_{L\in B_{ij}^c} e^{2\beta_t (c-1)\epsilon(t)} e^{-\beta_t s_L^*} (s_L^*)^2. \qquad (64)
\end{aligned}$$

By the induction assumption $\epsilon(t) < 1/2(c-1)\beta_t$ and then using the definition of λ, we have

$$\epsilon(t+1)^2 \leq \frac{e \sum_{L\in B_{ij}^c} e^{-\beta_t s_L^*} (s_L^*)^2}{|G_{ij}|} \leq \frac{e\lambda}{(1-\lambda)|B_{ij}|} \sum_{L\in B_{ij}^c} e^{-\beta_t s_L^*} (s_L^*)^2. \qquad (65)$$

Combining the lemma assumptions and the definition of M we have

$$\epsilon(t+1)^2 \leq \frac{e\lambda}{M(1-\lambda)\beta_t^2} = (\frac{r}{2(c-1)\beta_t})^2. \qquad (66)$$

Therefore the lemma is proved by taking the square root of both sides. $\square$

Proof of Lemma C.9. To prove this lemma, we first prove an upper bound on the suprema of weakly dependent empirical processes. For an index set $\mathcal{A}$ and corresponding random variables $\{X_\alpha\}_{\alpha\in\mathcal{A}}$, we make the following definitions:

- A subset $\mathcal{A}'$ of $\mathcal{A}$ is independent if $\{X_\alpha\}_{\alpha\in\mathcal{A}'}$ is independent.
- A family of pairs $(\mathcal{A}_k, w_k)$ is a fractional cover of $\mathcal{A}$ if $\sum_k w_k 1_{\mathcal{A}_k} \geq 1_{\mathcal{A}}$.
- A fractional cover $(\mathcal{A}_k, w_k)$ is proper if each set $\mathcal{A}_k$ is independent.

Lemma D.4. Assume $\{X_\alpha\}_{\alpha\in I}$ are identically distributed according to P. Assume $\mathcal{F}$ is a countable set of functions that are all P-measurable and for all $f \in \mathcal{F}$, $\|f\|_\infty \leq 1$. Let $Z = \sup_{f\in\mathcal{F}} |\sum_{\alpha\in I} f(X_\alpha)|$. Assume I admits a proper fractional cover $\{(I_j, w_j)\}_{j\in J}$, and $Z_j = \sup_{f\in\mathcal{F}} |\sum_{\alpha\in I_j} f(X_\alpha)|$. Let $\{p_j\}_{j\in J}$ be positive numbers such that $\sum_j p_j = 1$. Then

$$P(Z > \sum_j w_j \mathbb{E} Z_j + t) < \exp(-v\varphi(\frac{t}{Wv})) \qquad (67)$$

where $v = 2\min_j \mathbb{E} Z_j + \sup_{f\in\mathcal{F}} \text{Var}(f(X_\alpha))$ and $W = \sum_j w_j$.

Proof. We follow the proof strategy of [23]. By lemma 3.2 in [23] we can assume (I_j, w_j) is an exact fractional cover of I. We have

$$Z = \sup_{f \in \mathcal{F}} |\sum_{\alpha \in I} f(X_\alpha)| \tag{68}$$

$$\leq \sup_{f \in \mathcal{F}} |\sum_{\alpha \in I} \sum_j w_j 1_{I_j}(\alpha) f(X_\alpha)| \tag{69}$$

$$= \sup_{f \in \mathcal{F}} |\sum_j w_j \sum_{\alpha \in I} 1_{I_j}(\alpha) f(X_\alpha)| \tag{70}$$

$$= \sup_{f \in \mathcal{F}} |\sum_j w_j \sum_{\alpha \in I_j} f(X_\alpha)| \tag{71}$$

$$\leq \sum_j w_j \sup_{f \in \mathcal{F}} |\sum_{\alpha \in I_j} f(X_\alpha)| = \sum_j w_j Z_j. \tag{72}$$

Let p_j be any positive numbers such that $\sum_j p_j = 1$. By Jensen's inequality, for any $u > 0$,

$$\exp(u(Z - \sum_j \mathbb{E} Z_j)) \leq \exp(\sum_j p_j \frac{u w_j}{p_j}(Z_j - \mathbb{E} Z_j)) \leq \sum_j p_j \exp(\frac{u w_j}{p_j}(Z_j - \mathbb{E} Z_j)). \tag{73}$$

Since Z_j is the supremum of a sum of independent random variables, by theorem 2.1 in [3] we have

$$\mathbb{E}\exp(\frac{u w_j}{p_j}(Z_j - \mathbb{E} Z_j)) \leq \exp(\psi(-\frac{u w_j}{p_j}) v_j) \tag{74}$$

where $\psi(x) = e^{-x} - 1 + x$ and $v_j = 2\mathbb{E} Z_j + \sup_{f \in \mathcal{F}} \mathrm{Var}(f(X_\alpha))$. Let $p_j = w_j / W$. By definition of v, $v = \min_j v_j$. By Markov's inequality we have

$$P(Z - \sum_j \mathbb{E} Z_j \geq t) \leq e^{-ut} \mathbb{E} e^{u(Z - \sum_j \mathbb{E} Z_j)} \tag{75}$$

$$\leq e^{-ut} \frac{\sum_j w_j e^{\psi(-uW) v_j}}{W} \tag{76}$$

$$\leq e^{-ut} \frac{\sum_j w_j e^{\psi(-uW) v}}{W} \tag{77}$$

$$= e^{-ut + \psi(-uW) v} \tag{78}$$

$$= e^{-ut + (e^{uW} - 1 - uW) v}. \tag{79}$$

Taking the minimum of the right hand side with respect to u gives $P(Z \geq t) \leq e^{-v\varphi(t/Wv)}$. □

Now let's prove Lemma C.9. We slightly abuse the notation for simplicity. Throughout this proof we use B_{ij} as the set of all bad ij,c-paths. To use Lemma D.4, we need to construct a proper fractional cover of B_{ij}^c. Let $\Delta_1 = \lfloor |B_{ij}^c| / cm_{c-1} \rfloor$. Note that by the regular E-R condition, we know that each $L \in B_{ij}^c$ has at most cm_{c-1} cycles that are correlated with L. By Hajnal-Szemerédi theorem, there exists a partition of B_{ij}^c, namely $\{B_{ij,k}^c\}_{k=1}^{cm_{c-1}}$, where for any k, $|B_{ij,k}^c| = \Delta_1$ or $\Delta_1 + 1$, and all paths in $B_{ij,k}^c$ are independent. This induces a proper fractional cover $(B_{ij,k}^c, 1)$. By Lemma D.4, for any $t > 0$ we have

$$P(\sup_{f_\tau \in \mathcal{F}(\beta)} \sum_{L \in B_{ij}^c} f_\tau(s_L^*) > t + cm_{c-1} \max_k \mathbb{E} Z_k) < \exp(-v\varphi(\frac{t}{cm_{c-1} v})). \tag{80}$$

where $v = 2\min_k \mathbb{E} Z_k + V(\beta)$.

By lemma 7 of [25] we know that $\mathbb{E} Z_k \leq C_1 \sqrt{\log |B_{ij,k}^c| / |B_{ij,k}^c|}$. By $|B_{ij,k}^c| \geq \Delta_1$ we know $\log |B_{ij,k}^c| / |B_{ij,k}^c| \leq \log \Delta_1 / \Delta_1$.

By $\varphi(x) > \frac{x}{2}\ln(1+x)$ and the definition of Δ_1, let $t = |B_{ij}^c|(2C_1\sqrt{\log\Delta_1/\Delta_1} + V(\beta))$ in (80), we have

$$
\begin{aligned}
&P\left(\sup_{f_\tau \in \mathcal{F}(\beta)} \frac{1}{|B_{ij}^c|}\sum_{L\in B_{ij}^c} f_\tau(s_L^*) > V(\beta) + (2C_1 + \frac{1}{\Delta_1})\sqrt{\frac{\log\Delta_1}{\Delta_1}}\right)\\
&< \exp\left(-\frac{\ln 2}{2}\Delta_1(2C_1\sqrt{\frac{\log\Delta_1}{\Delta_1}} + V(\beta))\right).
\end{aligned} \tag{81}
$$

By the definition of m_{c-1} we know that $cm_{c-1} \sim \max(n^{c-3}p^{c-2}, n^\epsilon)$. Therefore $\Delta_1 = \Omega(\min(np, n^{c-2-\epsilon}p^{c-1}))$. Since $\Delta_1 \geq 1$, Lemma C.9 is proved by letting $K'' = 2C_1 + 1$. □

E. Extension to any linear group with the metric induced by the Frobenius norm

Our algorithm LongSync can be extended to any linear group with the metric induced by the Frobenius norm. Let $\mathcal{D}_\mathcal{G}(\boldsymbol{G}_1, \boldsymbol{G}_2) = \|\boldsymbol{G}_1 - \boldsymbol{G}_2\|_F$ be such metric defined on a linear group $\mathcal{G}$. The update rule of LongSync becomes:

$$
\begin{aligned}
s_{ij}^{(t)} &= \Big(\sum_{L\in N_{ij}^c} w_L^{(t)} d_L^2 / z_{ij}^{(t)}\Big)^{1/2}\\
&= \Big(\sum_{L\in N_{ij}^c} w_L^{(t)} \mathcal{D}_\mathcal{G}^2(\boldsymbol{G}_L, \boldsymbol{G}_{ij}) / z_{ij}^{(t)}\Big)^{1/2}\\
&= \left(\Big(\sum_{L\in N_{ij}^c} w_L^{(t)} \|\boldsymbol{G}_L - \boldsymbol{G}_{ij}\|_F^2\Big) / z_{ij}^{(t)}\right)^{1/2}\\
&= \left(\Big(\Big\langle \sum_{L\in N_{ij}^c}\sqrt{w_L^{(t)}}\boldsymbol{G}_L, \sum_{L\in N_{ij}^c}\sqrt{w_L^{(t)}}\boldsymbol{G}_L\Big\rangle - 2\Big\langle \sum_{L\in N_{ij}^c} w_L^{(t)}\boldsymbol{G}_L, \boldsymbol{G}_{ij}\Big\rangle + \sum_{L\in N_{ij}^c} w_L^{(t)}\Big\langle \boldsymbol{G}_{ij}, \boldsymbol{G}_{ij}\Big\rangle\Big) / \sum_{L\in N_{ij}^c} w_L^{(t)}\right)^{1/2}.
\end{aligned} \tag{82}
$$

With the same f_c and g_c in 3.1, we have the following proposition:

Proposition E.1. *The update rule of of LongSync for any linear group in equation (82) is equivalent to the following matrix operations:*

$$
\boldsymbol{S}^{(t)} = \left(\left(\Big\langle g_c(\sqrt{\boldsymbol{W}^{(t)}}, \boldsymbol{G}), g_c(\sqrt{\boldsymbol{W}^{(t)}}, \boldsymbol{G})\Big\rangle_{block} - 2\Big\langle g_c(\boldsymbol{W}^{(t)}, \boldsymbol{G}), \boldsymbol{G}\Big\rangle_{block}\right) \oslash f_c(\boldsymbol{W}^{(t)}) + \langle \boldsymbol{G}, \boldsymbol{G}\rangle_{block}\right)^{\odot 1/2} \tag{83}
$$

where $\boldsymbol{W}^{(t+1)} = \boldsymbol{A} \odot \exp(-\beta_t \boldsymbol{S}^{(t)})$.

Proof. We prove the proposition by comparing the ij-th element of the right hand side of equation (83) with (82). By the definition of blockwise inner product, the ij-th block of the right hand side of equation (83) is

$$
\left(\left(\Big\langle g_c(\sqrt{\boldsymbol{W}^{(t)}}, \boldsymbol{G})(i,j), g_c(\sqrt{\boldsymbol{W}^{(t)}}, \boldsymbol{G})(i,j)\Big\rangle - 2\Big\langle g_c(\boldsymbol{W}^{(t)}, \boldsymbol{G}), \boldsymbol{G}_{ij}\Big\rangle\right) / f_c(\boldsymbol{W}^{(t)}) + \langle \boldsymbol{G}_{ij}, \boldsymbol{G}_{ij}\rangle\right)^{1/2}.
$$

Note that by definition of g_c, $g_c(\sqrt{\boldsymbol{W}(t)}, \boldsymbol{G})(i,j) = \sum_{L\in N_{ij}^c}\sqrt{w_L^{(t)}}\boldsymbol{G}_L$, and $g_c(\boldsymbol{W}(t), \boldsymbol{G})(i,j) = \sum_{L\in N_{ij}^c} w_L^{(t)}\boldsymbol{G}_L$. By the definition of f_c, $f_c(\boldsymbol{W}^{(t)})(i,j) = \sum_{L\in N_{ij}^c} w_L^{(t)}$. By directly comparing the terms we know that the right hand side of equation (83) is the same as (82). □

In view of this vectorized update rule, we propose the vectorized LongSync iterations for any linear group with l_2 metric in algorithm 2.

We remark that the theory of LongSync can also be adapted as long as the group is 'well-conditioned', i.e. there exists constants $M_\mathcal{G}$ and $m_\mathcal{G}$ only depending on $\mathcal{G}$ such that for any $\boldsymbol{G} \in \mathcal{G}$, the absolute value of the eigenvalues of $\boldsymbol{G}$ is between $m_\mathcal{G}$ and $M_\mathcal{G}$.

Algorithm 2 (LongSync for any linear group)

Input: pairwise measurement matrix $\boldsymbol{G}$, adjacency matrix $\boldsymbol{A} \in [0,1]^{n \times n}$, cycle length c, positive parameters $\{\beta_t\}_{t \geq 1}$, time step T

$\boldsymbol{W}^{(0)}(i,j) \leftarrow \boldsymbol{A}$

for $t = 0 : T$ **do**

$$\boldsymbol{S}^{(t)} \leftarrow \left(\left(\left\langle g_c(\sqrt{\boldsymbol{W}^{(t)}}, \boldsymbol{G}), g_c(\sqrt{\boldsymbol{W}^{(t)}}, \boldsymbol{G}) \right\rangle_{\text{block}} - 2 \left\langle g_c(\boldsymbol{W}^{(t)}, \boldsymbol{G}), \boldsymbol{G} \right\rangle_{\text{block}} \right) \oslash f_c(\boldsymbol{W}^{(t)}) + \langle \boldsymbol{G}, \boldsymbol{G} \rangle_{\text{block}} \right)^{\odot 1/2}$$

$$\boldsymbol{W}^{(t+1)} \leftarrow \boldsymbol{A} \odot \exp(-\beta_t \boldsymbol{S}^{(t)})$$

end for

Output: edge weights $\boldsymbol{W}^{(T+1)}$, corruption levels $\boldsymbol{S}^{(T)}$