# Technical Report for the paper De-anonymizability of Social Network: Through the Lens of Symmetry

## A PROOF DETAILS IN SECTION 4.1

### A.1 Proof of Lemma ?? and Lemma ??

We first introduce the product of two permutations [? ].

*Definition A.1.* If $\sigma_1, \sigma_2$ are two permutations on $V$, then $\sigma_p$ is the product of $\sigma_1$ and $\sigma_2$, if $\sigma_p(v) = \sigma_2(\sigma_1(v))$, for any $v \in V$.

Clearly the product of two permutations (on the same domain) is also a permutation.

A proposition about the transitivity property of automorphism follows.

Proposition A.2. *If two permutations $\sigma_1, \sigma_2$ are both automorphisms on $G$, then the product of $\sigma_1$ and $\sigma_2$, $\sigma_p$, is also an automorphism on $G = (V, E)$.*

Proof. First, $\sigma_p$ is also a permutation. Since $\sigma_1$, $\sigma_2$ are both automorphisms of $G$, by definition

$$\forall(v_i, v_j) \in E \leftrightarrow (\sigma_1(v_i), \sigma_1(v_j)) \in E$$
$$\forall(v_i, v_j) \in E \leftrightarrow (\sigma_2(v_i), \sigma_2(v_j)) \in E$$

$\sigma_p$, the product of $\sigma_1, \sigma_2$, is a permutation that maps any $v$ to $\sigma_2(\sigma_1(v))$ in node set. And by the two formula above, we can see

$$\forall(v_i, v_j) \in E \leftrightarrow \sigma_2(\sigma_1(v_i)), \sigma_2(\sigma_1(v_j))) \in E$$

It shows that $\sigma_p$ is also an automorphism on $G$. □

Now comes the proof of Lemma **??**.

Proof of Lemma **??**. Clearly for each node $v$ there exists an orbit containing $v$, for an nontrivial automorphism (which maps each node to itself) will map $v$ to itself. Then the only thing we need to prove is that any two orbits are disjointed.

Actually that is easy to prove. Suppose there exists two different orbits $\mathscr{O}_1$, $\mathscr{O}_2$ containing a common node $v$, we try to yield a contradiction. Consider an arbitrary node $v_1$ in $\mathscr{O}_1$, and $v_2$ in $\mathscr{O}_2$. Since $v$ and $v_1$ are in the same orbit, there exists an automorphism $\sigma_1$ such that $v = \sigma_1(v_1)$. For the same reason, we have an automorphism $\sigma_2$ such that $v_2 = \sigma_2(v)$.

Then we denote $\sigma_p$ as the product of $\sigma_1$ and $\sigma_2$, that is, for each node $v$, $\sigma_p(v) = \sigma_2(\sigma_1(v))$. By the theorem above, $\sigma_p$ is also an automorphism of $G$. Notice $\sigma_p(v_1) = \sigma_2(\sigma_1(v_1)) = v_2$, which means shows $\sigma_p$ will map $v_2$ to $v_1$. Since our $v_1$ and $v_2$ are chosen arbitrarily, we can deduce that $v_1 \in \mathscr{O}_2$ and $v_2 \in \mathscr{O}_1$. Again, since the arbitrary choice of $v_1$ and $v_2$, we can claim that $\mathscr{O}_1$ and $\mathscr{O}_2$ contains the same nodes. That means they are actually the same orbit. That completes the proof of disjoint property. □

The proof of Lemma **??** is a straightforward one. Having known all the orbits of a graph, constructing an automorphism of a graph can be seen as deciding a permutation within each orbit, and then combining them. There are $|Orb_i|!$ permutations within the orbit $i$. Applying the multiplication principle, we can obtain that $|Aut(G)| = \prod_{i=1}^{k} |\mathscr{O}_i|!$.

### A.2 Proof of Theorem ??

We only need to prove that, for each element $M_{ij}$ in matching probability matrix $M$, we have $M_{ij} = \frac{1}{|Orb(i)|}$. By definition

$$\begin{aligned}
M_{ij} &= \sum_{\pi \in \Pi} P(\sigma_0 = \pi|\theta) \mathbb{1}\{\pi(i) = j\} \\
&\overset{(0)}{=} \sum_{\pi \in Aut(G)} \frac{1}{|Aut(G)|} \mathbb{1}\{\pi(i) = j\} \\
&= \frac{1}{|Aut(G)|} \sum_{\pi \in Aut(G)} \mathbb{1}\{\pi(i) = j\} \\
&\overset{(1)}{=} \frac{1}{|Aut(G)|} \frac{|Aut(G)|}{|Orb(i)|} \\
&= \frac{1}{|Orb(i)|}
\end{aligned}$$

In this formula, (0) holds due to the probability distribution we have obtained in Proposition **??**. For equality (1) it is equivalent to count the number of distinct automorphisms which mapped $i$ to $j$. Suppose $i$ and $j$ are in the orbit $\mathscr{O}_q$, then the number is equal to $\prod_{i=1}^{k, i \neq q} |\mathscr{O}_i|! * (|\mathscr{O}_q| - 1)!$, which is equal to $\frac{|Aut(G)|}{|Orb(i)|}$ according to Lemma **??**.

## B PROOF DETAILS IN SECTION ??

### B.1 Proof of Proposition ??

Proof. Determining which graph to be the underlying network is like an inferencing process, so it is reasonable to use Bayes' Rule to deal with the probability. By Bayes' Rule we can write the probability of $G$ given $G_1$,$G_2$ as:

$$P(G|G_1, G_2) = \frac{P(G_1, G_2|G)P(G)}{P(G_1, G_2)}$$

On the right side, (a) $P(G)$ is a constant since we have no preference on underlying network; (b) $P(G_1, G_2)$ is a normalized factor and is identical among different $G$. Thus we have

$$\begin{aligned}
P(G|G_1, G_2) &\propto P(G_1, G_2|G) \\
&\overset{(0)}{=} P(G_1|G)(G_2|G) \\
&\overset{(1)}{=} \hom(G_1, G) s_1^{|E_1|} (1 - s_1)^{|E| - |E_1|} \\
&\quad \hom(G_2, G) s_2^{|E_2|} (1 - s_2)^{|E| - |E_2|} \\
&\propto \hom(G_1, G) \hom(G_2, G) ((1 - s_1)(1 - s_2))^{|E|}
\end{aligned}$$

In formula, (0) holds because $G_1$ and $G_2$ are independent samplings from $G$. (1) holds because there are $\hom(G_i, G)$ ways to sample $G$ to get $G_i$. □

### B.2 Proof of Proposition ??

We only prove the proposition for $G_1$. For $G_2$ the proof is the same.

Given $G_1$ and $G$, if a permutation $f_i$ is proved to be the true mapping $f_0$, then: (1) $f_i$ is a (graph bijective) homomorphism from $G_1$ to $G$ (otherwise $G_1$ cannot be sampled from $G$) (2) In the sampling process, for the edges in $G$, all the edges that exist in $G_1$ are 'sampled in', while all other edges that are absent from $G_1$ are 'sampled out'.

By Bayes' Law we can write that for any homomorphism $f_i$ from $G_1$ to $G$, the probability that $f$ is the true mapping $f_0$ is:

$$P(f_0 = f_i|G_1, G) = \frac{P(G_1|f_0 = f_i, G)P(f_0 = f_i|G)}{P(G_1|G)}$$

here, (a) $P(f_0 = f_i|G)$ is a constant since we have no preference for any specific mapping; (b) $P(G_1|G)$ is a normalized factor and is identical among different $f_i$; (c) $P(G_1|f_0 = f_i, G) = (1 - s_1)^{|E|-|E_1|}s_1^{|E_1|}$ is constant since the edge number of $G$ and both $G_i$ are determined.

Therefore, each homomorphism $f_i$ has the same probability $\frac{1}{\hom(G_1,G)}$ to be the true mapping when $G$ is given. Similar for $G_2$.

## B.3 Proof of Proposition ??

We only prove the result of $G_1$, i.e. $C_G = A_1C_G$. The proof for $G_2$ is completely the same. Let $C' = A_1C_G$. Then

$$\begin{aligned} C'_{ij} &= \sum_{k\in V}(A_1)_{ik}(C_G)_{kj} \\ &= \sum_{k\in V}^{k\in Orb_{G_1}(i)} \frac{1}{|Orb_{G_1}(i)|}C_{kj} \end{aligned}$$

Here $Orb_{G_1}(i)$ represents the orbit in $G_1$ that contains i. We can see from the formula that the theorem holds if $(C_G)_{kj} = (C_G)_{ij}$ for any $k \in Orb_{G_1}(i)$. In fact, the latter can be proved as follows:

Suppose $i$ and $k$ are in the same orbit (of $G_1$). That indicates that there exists an automorphism $f$ (on $G_1$) that maps $i$ to $k$ ($f(i) = k$).

Then for each homomorphism $\sigma$ from $i$ (in $G_1$) to $j$ (in $G$), there exists a permutation $\sigma' = f \circ \sigma$. On one hand, $\sigma'$ is a homomorphism, since $G_1$ keeps invariant under the action of $f$. On the other hand, $\sigma'$ maps $k$ to $j$. This suggests that for each homomorphism that maps $i$(in $G_1$) to $j$(in $G$), there also exists a homomorphism mapping $k$(in $G_1$) to $j$(in $G$), and vice versa. Therefore, the number of homomorphisms that map $i$ to $j$ and map $k$ to $j$ is equal.

Recall that $(C_G)_{ij} = \frac{1}{k_1}c_{ij}$, which is determined by the number of homomorphisms that match $i$ (in $G_1$) to $j$ (in $G$). Thus $(C_G)_{ij} = (C_G)_{kj}$ for any $i$,$k$ in the same orbit. Therefore, $C_G = A_1C_G$. Similarly $D_G = A_2D_G$.

## B.4 Proof of Theorem ??

We denote the matrix calculated by Equation ?? as $M'$, and the matching probability matrix calculated via the original method as $M$. Then we need to prove that $M' = M$. First, it is easy to see that $M'$ is also a doubly stochastic matrix. Therefore, we only need to prove that, the corresponding elements in $M$ and $M'$ are in proportion. Starting from Equation ??, we have

$$\begin{aligned} M_{ij} &= \sum_{G\in\mathcal{G}}\sum_{\pi\in\Pi} P(G|\theta)P(\sigma_0 = \pi|\theta, G)\mathbb{1}\{\pi(i) = j\} \\ &= \sum_{\pi\in\Pi}\sum_{G\in\mathcal{G}} P(G|\theta)P(\sigma_0 = \pi|\theta, G)\mathbb{1}\{\pi(i) = j\} \\ &\overset{(0)}{=} \sum_{\pi\in\Pi}\sum_{G\in\mathcal{G}}\sum_{f\in\Pi} P(G|\theta)P(f|\theta, G)P(h|\theta, G))\mathbb{1}\{\pi(i) = j\} \\ &\propto \sum_{\pi\in\Pi}\sum_{G\in\mathcal{G}}\sum_{f\in\Pi} (\bar{s_1}\bar{s_2})^{|E|}\hom(G_1, G)\hom(G_2, G) \\ &\quad P(f|\theta, G)P(h|\theta, G)\mathbb{1}\{\pi(i) = j\} \\ &\propto \sum_{\pi\in\Pi}\mathbb{1}\{\pi(i) = j\}\sum_{G\in\mathcal{G}}(\bar{s_1}\bar{s_2})^{|E|}\sum_{f\in\Pi}\hom(G_1, G)\hom(G_2, G) \\ &\quad P(f|\theta, G)P(h|\theta, G) \\ &= \sum_{\pi\in\Pi}\mathbb{1}\{\pi(i) = j\} \\ &\quad \sum_{G\in\mathcal{G}}(\bar{s_1}\bar{s_2})^{|E|}\mathbb{1}\{\exists f : f(G_1)\subset G\}\mathbb{1}\{h(G_2)\subset G\} \end{aligned}$$

In this formula, $h = \pi^{-1} \circ f$ so that we can have $f \circ h^{-1} = \pi$.

Therefore, for each permutation $\pi$, we count up all the graphs $G$ whose edge set is a superset of both that of $G_1$ and $G_2$. Obviously, the condition holds iff $G$ is the spanning super graph of the union of $G_1$ and $G_2$ under permutation $\pi$. We denote the union of $G_1$ and $G_2$ under permutation $\pi$ as $G_\pi = (V_\pi, E_\pi)$, and denote $N = \binom{n}{2}$. We have,

$$\begin{aligned} M_{ij} &\propto \sum_{\pi\in\Pi}\mathbb{1}\{\pi(i) = j\} \\ &\quad \sum_{i=0}^{N-|E_\pi|}\left(\binom{N-|E_\pi|}{i}\bar{s_1}^{|E_\pi|-|E_1|+i}\bar{s_2}^{|E_\pi|-|E_2|+i}\right) \\ &= \sum_{\pi\in\Pi}\mathbb{1}\{\pi(i) = j\}\bar{s_1}^{|E_\pi|-|E_1|}\bar{s_2}^{|E_\pi|-|E_2|} \\ &\quad (1 + (\bar{s_1}\bar{s_2}))^{N-|E_\pi|} \\ &\propto \sum_{\pi\in\Pi}\mathbb{1}\{\pi(i) = j\}\bar{s_1}^{|E_\pi|}\bar{s_2}^{|E_\pi|}(1 + (\bar{s_1}\bar{s_2}))^{-|E_\pi|} \\ &= \sum_{\pi\in\Pi}\mathbb{1}\{\pi(i) = j\}\frac{\bar{s_1}\bar{s_2}}{1 + \bar{s_1}\bar{s_2}}^{|E_\pi|} \propto M'_{ij} \end{aligned}$$

which completes the proof.

# C A CASE STUDY : DE-ANONYMIZABILITY IN ERDŐS-RÈNYI RANDOM GRAPHS

This part provides the technique details in the case study of de-anonymizability in Erdős-Rènyi graph.

As [**?** ] has mentioned, the symmetric structure in real-world network is more likely to be 'local'. The following graph can serve as an example to show the difference between an artificially constructed graph and a network-generated graph. The left one is the famous (artificially constructed) Petersen Graph, of which the symmetry structure is completely global(that is, we have to search the whole graph to assert the existence of automorphism in the graph). The right one is an instance generated by Erdős-Rènyi model G(n=10,

p=0.2). The only symmetric node pair here is (15,18) and (19,20), and detecting both of them only requires local information. On the other hand, in instances of network model, large, complicated symmetric structures are unlikely to appear.

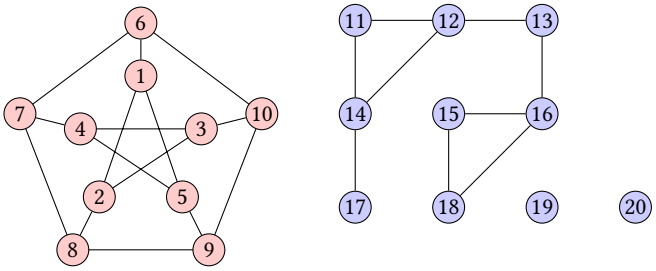


**Figure 1: An comparison. The left is the (artificially constructed) Peterson Graph, in which detecting the automorphism requires searching all the nodes. The right is an instance from G(10,0.2). Only two node pairs, (15,18) and (19,20), are symmetric here.**

From this phenomenon, we calculate the de-anonymizability via detecting some local automorphism structures in a network. This method is somehow like the *motif detection* [**?** ]. We will study the fully sampled case in detail, and then briefly demonstrate how we deal with the partially sampled case based on the result from the fully sample case.

## C.1 Fully Sampled Case

In fully sampled cases we only need to count the orbits of a graph to get its de-anonymizability. As we have said, we only focus on those local symmetric structures of a graph. Slightly abusing the concept, we also use *motif* to refer to those local subgraphs or patterns. In this section, a motif is merely a set of nodes with some specific pattern (especially the pattern in symmetry). In the analysis of Erdős-Rènyi graph, we will clearly specify which kind of patterns we consider. For a graph $G = (V, E)$, a motif is denoted by $V_s \subset V$, and we define $T(V_s) = |V_s| - |Orb(V_s)|$, where $|Orb(V_s)|$ means the number of orbits in the graph $G$ that contains all the nodes in $V_s$. Notice that, $T(V_s)$ means the number of orbits that $V_s$ 'contracts'. The following theorem demonstrates an upper bound of the orbit

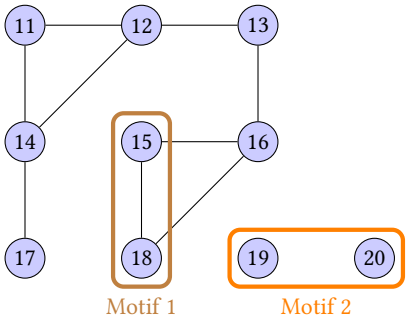


**Figure 2: The instance from G(10,0.2). This picture is to show what a motif is.**

number in a graph , which can be used to show the approximate de-anonymizability of a graph.

Theorem C.1. *Given a graph $G = (V, E)$, suppose we have $k$ motifs of $G$ $V_{s1}, V_{s2}, \ldots, V_{sk}$. If any two motifs are vertex-disjoint, then we have $|Orb(G)| \leq |N| - \sum_{i=1}^{k} T(V_{si})$. The equality holds when:*

- *There does not exist a symmetric node pair across two motifs. That is, for any $V_{si}$ and $V_{sj} (i \neq j)$, for any $v_i \in V_{si}$ and $v_j \in V_{sj}$, no automorphism of $G$ will map $v_i$ to $v_j$.*
- *All the motifs are found. That is, for any node $v$ that does not belong to any of $V_{s1}, V_{s2}, \ldots, V_{sk}$, all the automorphism of $G$ maps $v$ to itself.*

Proof. The inequality is easy to prove by the definition of orbit. For the nodes that does not belong to any motif, it owns at most one orbit. Thus

$$
\begin{aligned}
|Orb(G)| &\leq |Orb(V - \bigcap_{i=1}^{k} V_{si})| + \sum_{i=1}^{k} |Orb(V_{si})| \\
&\leq |N| - \sum_{i=1}^{k} |V_{si}| + \sum_{i=1}^{k} |Orb(V_{si})| \\
&= |N| - \sum_{i=1}^{k} T(V_{si})
\end{aligned}
$$

The equality holds when no additional symmetric node pair exists, which can be expressed by the two given conditions. □

We apply this theorem into the context of Erdős-Rènyi graph. We only consider the situation when $p \leq 0.5$, for a graph and its complement have the same automorphisms. We count up two commonest kinds of motifs in a graph:

(1) The fruits in a cherry-like structure.
(2) The small connected components.

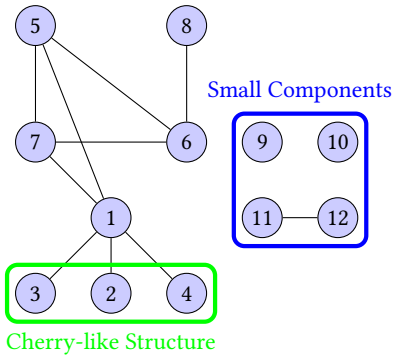


**Figure 3: Two kinds of motifs we consider here.**

In the sequel we discuss them respectively.

*C.1.1 Cherry-like Motif.* We first consider cherry-like motif, denoted as $S$. The cherry-like motif is shown above in Figure 3. Precisely, a set of $k$ nodes forms a '$k$-fruit cherry-like' motif(in short '$k$-fruit') *iff* (1) the degrees of all of them are equal to one (2) all of them are connected to the same node. For a graph $G$, we define

$S_k = \{V_{s_k}\}$ as the set of the all the '$k$-fruit' of the $G$. We can obtain the expectation of how many times these type of structures appear. Specifically, for $k$ nodes, the probability that all of them are merely connected to the same node is

$$P_{V_{s_k}} = (n-k)p^k(1-p)^{k(n-k-1)} * p^{\binom{k}{2}}$$

The expectation appearance of '$k$-fruit' over the whole graph is

$$E(|S_k|) = \binom{n}{k} P_{s_k} = \binom{n}{k}(n-k)p^k(1-p)^{k(n-k-1)} * p^{\binom{k}{2}}$$

And, since each '$k$-fruit' obtains only one orbit,

$$\sum_{V_{s_k} \in S_k} T(V_{s_k}) = (k-1)|S_k|$$

We define $S = \bigcup_{k=2}^{n} S_k$ as the set of all cherry-like motifs in graph $G$. We cannot simply add the result above, since repeated counting appears. For example, a '3-fruit' structure also contains $C_3^2 = 3$ '2-fruit' structures. By the inclusive-exclusive principle, we have

$$T(S) = \sum_{V_s \in S} T(V_s) = \sum_{k=2}^{n} (-1)^k \binom{n}{k}(n-k)p^k(1-p)^{k(n-k-1)} * p^{\binom{k}{2}}$$

by which we can calculate the (expectation) number of 'cherry-like' motifs and the number of orbits they contract given any specific parameters $n$, $p$.

*C.1.2 Small connected components.* According to [**?** ], it has been shown that in Erdős-Rènyi graph $G(n,p)$ with $np > \log(n)$, almost every graph generated by $G(n,p)$ is asymmetric. On the other hand, when $np \leq \log(n)$, most graph generated by $G(n,p)$ is not fully connected, and the existence of small connected components (with contrast to the giant component in Erdős-Rènyi graph) are likely to contain symmetric nodes. Therefore, we consider another type of motif: small connected components (denoted as $C$). Here we simply define that small connected components in a graph are all the connected components whose sizes are less than a threshold $k$. Notice that if two components share the same shape, not only the nodes within the motif may belong to the same motif (since most of the small components are symmetric), the corresponding nodes in different components will also belong to the same orbit. To reduce the number of components we have to enumerate, we use the fact that almost all the small components in Erdős-Rènyi graph are trees[**?** ]. Therefore, we only enumerate all the trees with size less than $k$. As a trade-off, a larger $k$ will detect more automorphisms and improve the accuracy the result, while exponentially increases the number of components that are necessary to enumerate. As an illustration, Figure 4 shows all 13 kinds of trees whose size is less than or equal to 6 [**?** ].

For each type, the expectation of times of its appearance is calculated in order to obtain the number of orbits it contracts. As an example, we show in detail how to calculate the expectation of $C_1$ (i.e. an isolated node). The probability that it is connected to none of the other points is

$$P(deg_i = 0) = (1-p)^{n-1}$$

Then the expectation over the whole graph is equal to

$$E(c_1) = n(1-p)^{n-1}$$

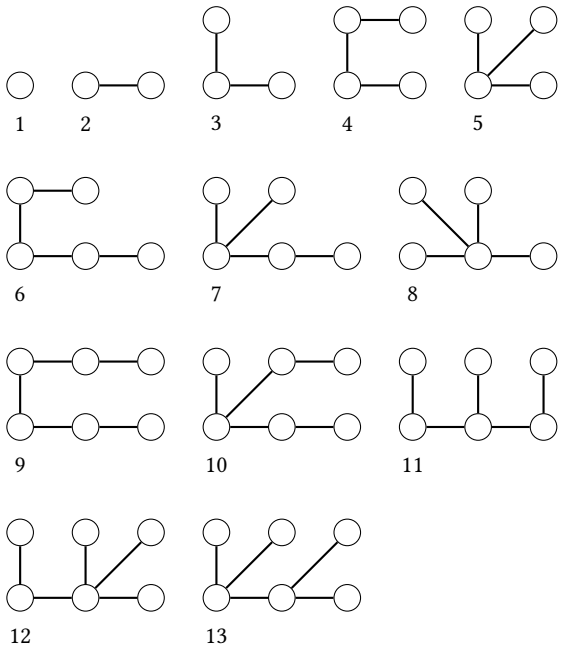


**Figure 4: Different types of small components.**

which shows the number of $C_1$ motifs (i.e. isolated nodes) in G (in expectation). For type 3,5,7,8,12,13, cherry-like structure exists in those components. Therefore, we only contract the orbit **across** the components. Take type 3 as an example, the expectation of appearance of type 3 components is:

$$E(c_3) = \binom{n}{3} \frac{3!}{Aut(c_3)} p^2(1-p)^{3(n-3)+1}$$

It is easy to observe that, the orbit number, after contraction in C, is 2. However, in each type-3 component (i.e. $C_3$), the two 'fruit' nodes have already been contracted into one orbit after the analysis Section C.1.1, and therefore, only $2 * E(c_3) - 2$ (rather than $3 * E(c_3) - 2$) orbits are contracted. We list the results of each type in the Table 1. Consequently, $T(C) = \sum_i T(c_i)$ represents the total number of orbits contracted by small connected components. In fact, a larger component size threshold $k$ can detect more small components, but when one motif does not exist, extra orbits are contracted wrongly. Therefore, we dynamically decide $k$ when conducting our experiments. After analyzing those two kinds of motifs, the de-anonymizability of a graph $G$ is then given by

$$E_{\sigma^*} = E[Orb(G)] = n - T(S) - T(C)$$

Notice that in the above formula, the right hand side is a expression of $n$ and $p$. Substituting $(n,p)$ in this formula with specified values, we can get an approximation of the de-anonymizability of $G$.

For ease of understanding, let us now take an ER graph with $n = 1000, p = 1/500$ as an example. After some calculation we can get

$$T(S) = 33.69, T(C) = 181.49$$
$$E_{\sigma^*} = n - T(S) - T(C) = 784.82$$

**Table 1: $T(C)$ for each kind of motifs**

| i | $E(c_i)$ | $T(c_i)$ |
|---|---|---|
| 1 | $n(1-p)^{n-1}$ | $2*E(c_1)-1$ |
| 2 | $\binom{n}{2}p(1-p)^{2(n-2)}$ | $2*E(c_2)-1$ |
| 3 | $\binom{n}{3}\frac{3!}{2}p^2(1-p)^{3(n-3)+1}$ | $2*E(c_3)-1$ |
| 4 | $\binom{n}{4}\frac{4!}{2}p^3(1-p)^{4(n-4)+3}$ | $2*E(c_4)-2$ |
| 5 | $\binom{n}{4}\frac{4!}{3}p^3(1-p)^{4(n-4)+3}$ | $4*E(c_5)-2$ |
| 6 | $\binom{n}{5}\frac{5!}{2}p^4(1-p)^{5(n-5)+6}$ | $5*E(c_6)-3$ |
| 7 | $\binom{n}{5}\frac{5!}{2}p^4(1-p)^{5(n-5)+6}$ | $4*E(c_7)-4$ |
| 8 | $\binom{n}{5}\frac{5!}{4!}p^4(1-p)^{5(n-5)+6}$ | $2*E(c_8)-1$ |
| 9 | $\binom{n}{6}\frac{6!}{2}p^5(1-p)^{6(n-6)+10}$ | $6*E(c_9)-2$ |
| 10 | $\binom{n}{6}\frac{6!}{2}p^5(1-p)^{6(n-6)+10}$ | $6*E(c_9)-4$ |
| 11 | $\binom{n}{6}\frac{6!}{2}p^5(1-p)^{6(n-6)+10}$ | $6*E(c_10)-4$ |
| 12 | $\binom{n}{6}\frac{6!}{3!}p^5(1-p)^{6(n-6)+10}$ | $4*E(c_11)-4$ |
| 13 | $\binom{n}{6}\frac{6!}{8}p^5(1-p)^{6(n-6)+10}$ | $4*E(c_12)-2$ |

[1] which means in a graph given by $G(1000, 1/500)$, more than 3/4 nodes can be (expected to be) de-anonymized. We will validate the result in the experiment section. The following theorem is used to prove the correctness of the method, which shows that, there is almost no symmetric node pairs in the giant component of Erdős-Rènyi graphs.

Theorem C.2. *For the Erdős-Rènyi model, $p = \Omega(\frac{1}{n})$ and $p \le \frac{1}{2}$, there are $o(1)$ nodes that are symmetric in the giant component.*

Proof. Suppose there are $n$ nodes in the giant component and the edge existence probability is $p$, where $p = \Omega(\frac{1}{n})$ and $p \le \frac{1}{2}$. The distribution of the node degree satisfies Poisson distribution. We aim to figure out the the probability of two nodes $N_a, N_b$ being symmetric given that both of their degrees are larger than one. Before the derivation, we first denote the nodes that are both $N_a$ and $N_b$'s neighbors to be 1-hop shared neighbors, and denote the nodes that are only $N_a$'s or $N_b$'s neighbors to be 1-hop free neighbors. If $N_a$ and $N_b$ are symmetric, these 1-hop free neighbors can be divided into several pairs, and each of them contains one $N_a$'s neighbor and one $N_b$'s, which are also symmetric. Similarly, we denote those nodes that connect two symmetric $(k-1)$-hop free neighbors as $k$-hop shared neighbors and nodes that only connect to one of them as $k$-hop free neighbors. Suppose the numbers of $N_a$'s and $N_b$'s $k$-hop free neighbors are $n_{ak}$ and $n_{bk}$, and the numbers of $N_a$'s and $N_b$'s $k$-hop share neighbors are $n'_{ak}$ and $n'_{bk}$.

If $N_a, N_b$ are symmetric, $n_{ak} = n_{bk} = n_k$ and $n'_{ak} = n'_{bk} = n'_k$. Moreover, since their $k$-hop free neighbors can be divided into symmetric pairs, the number of $N_a$'s $k$-hop free neighbors whose degrees are even is equal to that of $N_b$'s. The probability of a node with even degree is:

$$P_{even} = \sum_{j=1}^{\lfloor \frac{n-1}{2} \rfloor} \frac{\lambda^{2j}}{(2j)!} e^{-\lambda}, \tag{1}$$

[1] In calculation, we do some modification to our formula in order to prevent extra orbits to be contracted. The detail will be described in the experiment part.

where $\lambda = (n-1)p$. Since $n \to \infty$, we can derive that

$$\frac{P_{even}}{P_{odd}} = \frac{\sum_{j=1}^{\lfloor \frac{n-1}{2} \rfloor} \frac{\lambda^{2j}}{(2j)!} e^{-\lambda}}{\sum_{j=1}^{\lfloor \frac{n-2}{2} \rfloor} \frac{\lambda^{2j+1}}{(2j+1)!} e^{-\lambda}} \to \frac{e^{\lambda} + e^{-\lambda}}{e^{\lambda} - e^{-\lambda}} \to 1 \tag{2}$$

Therefore, $P_{even} \to \frac{1}{2}$ and $P_{odd} \to \frac{1}{2}$. Let $A_1$ be the event that the number of $N_a$'s $k$-hop free neighbors whose degrees are even is equal to that of $N_b$'s. Based on Stirling's formula, we can get that:

$$\begin{aligned} P_{A_1} &= \sum_{i=0}^{n-1} \left[ P_{even}^{i} P_{odd}^{n_k - i} \binom{n_k}{i} \right]^2 \\ &\to \sum_{i=0}^{n-1} 2^{-2n_k} \binom{n_k}{i}^2 \\ &= 2^{-2n_k} \binom{2n_k}{n_k} \\ &= \frac{(2n_k)!}{2^{2n_k} (n_k!)^2} \\ &\to \frac{\sqrt{4\pi n_k} \left(\frac{2n_k}{e}\right)^{2n_k}}{2\pi n_k 2^{2n_k} \left(\frac{n_k}{e}\right)^{2n_k}} \\ &= \frac{1}{\sqrt{\pi n_k}} \end{aligned} \tag{3}$$

Furthermore, when $A_1$ is satisfied, without loss of generality, we suppose the number of $k$-hop free neighbors with even degree is larger than or equal to $\frac{n_k}{2}$. Let $A_2$ be the event that the number of $N_a$'s $k$-hop free neighbors whose degrees satisfy $d = 4i$, where $i \ge 0$, is equal to that of $N_b$'s. With the same derivation, we can get that $P_{A_2} \ge \sqrt{\frac{2}{\pi n_k}}$. With the same analysis, we can also derive that $P_{A_3} \ge \sqrt{\frac{4}{\pi n_k}}$, $P_{A_4} \ge \sqrt{\frac{8}{\pi n_k}}$ and $P_{A_5} \ge \sqrt{\frac{16}{\pi n_k}}$, where $A_3, A_4$ and $A_5$ are events that the number of $N_a$'s $k$-hop free neighbors whose degrees satisfy $d = 8i$, $d = 16i$ and $d = 32i$, where $i \ge 0$, is equal to that of $N_b$'s respectively.

Suppose $N_a, N_b$ have $k$-hop free neighbors, where $k \le l$, then we can bound the probability that $N_a, N_b$ are symmetric by:

$$\begin{aligned} P_{(N_a,N_b)\in sym} &\le P_{deg(N_a)=deg(N_b)} \prod_{k=1}^{l} \prod_{j=1}^{5} P_{A_j} \\ &= P_{deg(N_a)=deg(N_b)} \prod_{k=1}^{l} \frac{32}{(\pi n_k)^{\frac{5}{2}}}, \end{aligned} \tag{4}$$

where $P_{deg(N_a)=deg(N_b)} = \sum_{i=2}^{n-1} \left(\frac{\lambda^i}{i!} e^{-\lambda}\right)^2$.

If there exists $k_0$ satisfies that $n_{k_0} = \Theta(n)$, then we have:

$$P_{(N_a,N_b)\in sym} \le \frac{32}{(\pi n_{k_0})^{\frac{5}{2}}} = o(\frac{1}{n^2}). \tag{5}$$

Then we have $\binom{n}{2} P_{(N_a,N_b)\in sym} = o(1)$.

Otherwise, for any constant $\epsilon > 0$ and two nodes $N_1, N_2$ in the network, let $B$ be the event: $deg(N_1) = deg(N_2) \ge (\frac{1}{2} + \epsilon)n$. Since

$p \le \frac{1}{2}$, then $\lambda = (n-1)p < \frac{n}{2}$ and we have:

$$\begin{aligned} P_B &= \sum_{i=(\frac{1}{2}+\epsilon)n}^{n-1} \left(\frac{\lambda^i}{i!}e^{-\lambda}\right)^2 \le \max_{i\ge(\frac{1}{2}+\epsilon)n}\left(\frac{\lambda^i}{i!}e^{-\lambda}\right)\sum_{i=(\frac{1}{2}+\epsilon)n}^{n-1}\frac{\lambda^i}{i!}e^{-\lambda} \\ &< \left(\frac{\lambda}{\frac{1+\epsilon}{2}n}\right)^{\frac{\epsilon}{2}n} e^{-\lambda} < \left(\frac{1}{1+\epsilon}\right)^{\frac{\epsilon}{2}n}. \end{aligned} \tag{6}$$

Then we have $\binom{n}{2}P_B = o(1)$, which means there are almost no node pairs that have the same degree larger than $(\frac{1}{2}+\epsilon)n$. Therefore, for $N_a, N_b$ or two nodes in k-hop free heighbors, suppose the degrees of them are both $d$, where $2 < d < (\frac{1}{2}+\epsilon)n$. The probability of the number of nodes that connect only one of them is $n_f$ satisfies $P_f \le \sum_{d=2}^{(\frac{1}{2}+\epsilon)n} P_{deg=d}\binom{d}{d-n_f}p^{d-n_f}$, when $n_f = o(d)$, we can derive that $P_f = O(\frac{1}{n^2})$. Therefore, $n_f = \Theta(d)$ with probability 1 and the total number of free neighbors is $\sum_{k=1}^{l} n_k = n-1-\sum_{k=1}^{l} n'_k = \Theta(n)$. Then we can derive that

$$\begin{aligned} P_{(N_a,N_b)\in sym} &< \prod_{k=1}^{l}\frac{32}{(\pi n_k)^{\frac{5}{2}}} \\ &\le \frac{32}{\pi^{\frac{5}{2}}(\sum_{k=1}^{l} n_k - l)^{\frac{5}{2}}}\prod_{k=1}^{l-1}\frac{1}{\sqrt{\pi}} \\ &= \frac{32}{\pi^{\frac{5}{2}}\left(\pi^{\frac{l-1}{5}}(\sum_{k=1}^{l} n_k - l)\right)^{\frac{5}{2}}} \\ &< \frac{32}{\pi^{\frac{5}{2}}\left(\sum_{k=1}^{l} n_k\right)^{\frac{5}{2}}} \\ &= o(\frac{1}{n^2}) \end{aligned} \tag{7}$$

Therefore, $\binom{n}{2}P_{(N_a,N_b)\in sym} = o(1)$.

Above all, there are almost no symmetric node pairs with degree larger than 1 in the network.

□

## C.2 Partially Sampled Case

The partially sampled case is relatively difficult to analyze. First, the assumption we made in Section **??** that each $G$ has the same prior probability does not hold, which brings difficult to find the probability distribution of the true mapping $\sigma_0$. Therefore, we bound the de-anonymizability of the problem using the de-anonymizability of $G_1$ and $G_2$ according to the result in **??**.

The underlying network $G$ is generated by model $G(n,p)$, and $G_1$, $G_2$ are independent samplings of $G$, the sampling rate of which is $s_1$,$s_2$, respectively. In fact, $G_1$ and $G_2$ can be seen generated by $G(n, ps_1)$ and $G(n, ps_2)$ respectively. We can calculate the orbit number of $G_1$ and $G_2$, and take the smaller one as an upper bound of de-anonymizability of the problem.