

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

This document is a technical report, which supports and provides technical details of the proof of key theorems of the paper ‘De-anonymizability of Social Network: Through the Lens of Symmetry’ (hereafter denoted as the *De-anonymizability* paper).

The symbols, models and denotations of this report are the same as the ‘De-anonymizability’ paper. To refer to the propositions, definitions, equations or theorems in the ‘De-anonymizability’ paper, the original index of the ‘De-anonymizability’ paper is used in this technical report.

1 PROOF OF PROPOSITION 4.1

PROPOSITION 1.1 (PROPOSITION 4.1). *Given parameter $\theta = (G_1, G_2, s_1, s_2)$ of a de-anonymization problem, for any graph $G = (V, E)$, a prior probability of G is given, denoted as $P(G)$. The probability of its being the ground truth underlying network is propositional to*

$$P(G) \text{hom}(G_1, G) \text{hom}(G_2, G) ((1 - s_1)(1 - s_2))^{|E|}$$

where $\text{hom}(F, G)$ is the bijective homomorphism counting from F to G . More precisely,

$$P(G|\theta) = HP(G) \text{hom}(G_1, G) \text{hom}(G_2, G) ((1 - s_1)(1 - s_2))^{|E|}$$

where $H = \sum_{G \in \mathcal{G}} P(G) \text{hom}(G_1, G) \text{hom}(G_2, G) ((1 - s_1)(1 - s_2))^{|E|}$ is a normalization parameter.

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, $P(G_1, G_2)$ is a normalized factor and is identical among different G . Thus we have

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

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

2 PROOF OF PROPOSITION 4.2

PROPOSITION 2.1 (PROPOSITION 4.2). *Given underlying network G and parameter θ of a de-anonymization problem, each homomorphism mapping f_i from G_1 to G has the probability of $\frac{1}{\text{hom}(G_1, G)}$ to be the true mapping from G_1 to G . Similarly, each h_i has the probability of $\frac{1}{\text{hom}(G_2, G)}$ to be the true mapping from G_2 to G .*

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 $\frac{1}{\text{hom}(G_1, G)}$ to be the true mapping when G is given. Similar for G_2 .

3 PROOF OF PROPOSITION 4.7

PROPOSITION 3.1 (PROPOSITION 4.7). *Given problem θ and an presumed underlying network G , let A_1, A_2 be their automorphism transition matrices of G_1, G_2 , respectively. Let C_G, D_G be the homomorphism transition matrix from G_1 and G_2 , respectively, to G . Then each homomorphism transition matrix keeps invariant under the multiplication of corresponding automorphism transition matrix. Precisely, $C_G = A_1 C_G, D_G = A_2 D_G$*

We only prove the result of G_1 , i.e. $C_G = A_1 C_G$. The proof for G_2 is completely the same. Let $C' = A_1 C_G$. Then

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

Here $\text{Orb}_{G_1}(i)$ 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 $k \in \text{Orb}_{G_1}(i)$. 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 σ from i (in G_1) to j (in G), there exists a permutation $\sigma' = f \circ \sigma$. On one hand, σ' is a homomorphism, since G_1 keeps invariant under the action of f . On the other hand, σ' 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 $(C_G)_{ij} = \frac{1}{k_i} c_{ij}$, 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, $C_G = A_1 C_G$. Similarly $D_G = A_2 D_G$.

4 PROOF OF THEOREM 5.2

THEOREM 4.1. *Given a de-anonymization problem with parameter $\theta = (G_1, G_2, s_1, s_2)$, let Π be all the permutations from V_1 to V_2 (i.e. all the permutations on V). Then matching probability matrix M can be calculated by*

$$M_{ij} = Z \sum_{\pi \in \Pi} \mathbb{1}\{\pi(i) = j\} \left(\frac{s_1 s_2}{1 + s_1 s_2} \right)^{|E_\pi|} \quad (1)$$

In this formula, $s_1 = 1 - s_1$, $s_2 = 1 - s_2$, $|E_\pi|$ is the edge number of the union of G_1 and G_2 under permutation π , Z is a normalization parameter which is identical among all elements in M . In particular

$$Z = \sum_{\pi \in \Pi} \left(\frac{s_1 s_2}{1 + s_1 s_2} \right)^{|E_\pi|}$$

We denote the matrix calculated by Equation 1 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 1, we have

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

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

Therefore, for each permutation π , 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 π . We denote the union of G_1 and G_2 under permutation π as $G_\pi = (V_\pi, E_\pi)$, and denote $N = \binom{n}{2}$. We

have,

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

which completes the proof.

5 A CASE STUDY : DE-ANONYMIZABILITY IN ERDŐS-RÉNYI RANDOM GRAPHS

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

As [2] 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-Rényi 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.

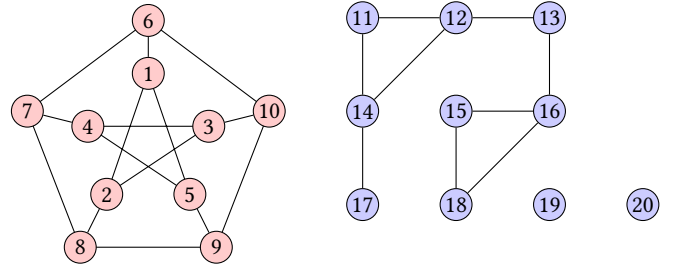


Figure 1: An comparison. The left is the (artificially constructed) Petersen Graph, in which detecting the automorphism requires searching all the nodes. The right is an instance from $G(10,0.2)$. Only two node pairs, (15,18) and (19,20), are symmetric here.

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* [3]. 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.

5.1 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-Rényi 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 $T(V_s) = |V_s| - |\text{Orb}(V_s)|$, where $|\text{Orb}(V_s)|$ 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

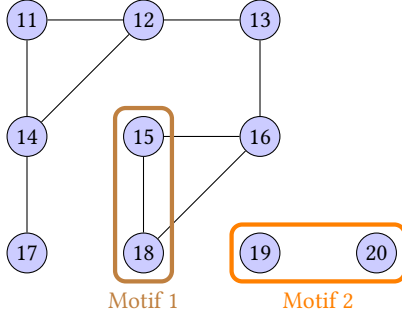


Figure 2: The instance from $G(10,0.2)$. This picture is to show what a motif is.

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

THEOREM 5.1. *Given a graph $G = (V, E)$, suppose we have k motifs of G $V_{s1}, V_{s2}, \dots, V_{sk}$. If any two motifs are vertex-disjoint, then we have $|\text{Orb}(G)| \leq |N| - \sum_{i=1}^k T(V_{si})$. 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 $V_{s1}, V_{s2}, \dots, V_{sk}$, 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

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

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

We apply this theorem into the context of Erdős-Rényi 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.

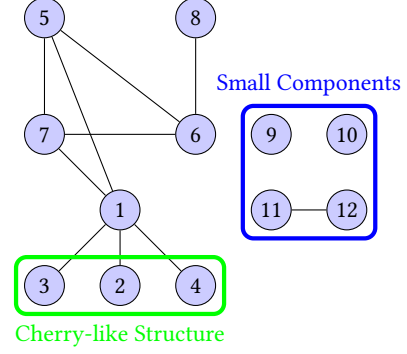


Figure 3: Two kinds of motifs we consider here.

In the sequel we discuss them respectively.

5.1.1 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 $S_k = \{V_{sk}\}$ as the set of 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

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

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,

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

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 ' k -fruit' motifs and the number of orbits they contract given any specific parameters n, p .

5.1.2 Small connected components. According to [1], it has been shown that in Erdős-Rényi 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-Rényi 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-Rényi graph are trees[1]. 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 [4].

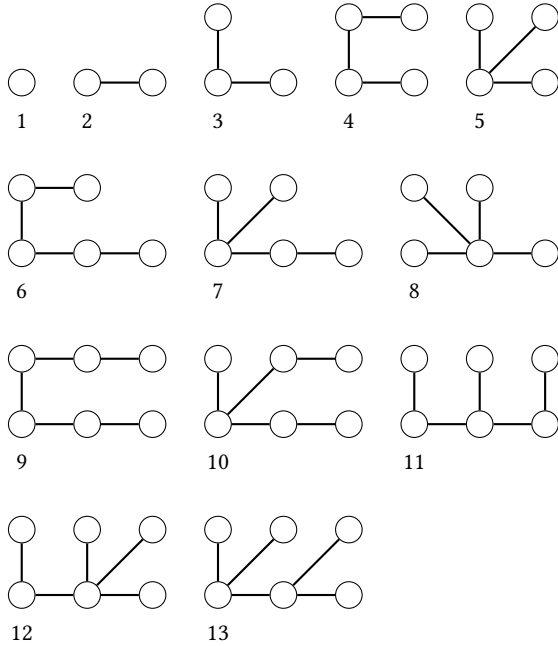


Figure 4: Different types of small components.

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

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

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 5.1.1, 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

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	$\binom{n}{6} \frac{6!}{8} p^5 (1 - p)^{6(n-6)+10}$	$4 * E(c_{12}) - 2$

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

$$T(S) = 33.69, T(C) = 181.49$$

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

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

¹In 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.

prove the correctness of the method, which shows that, there is almost no symmetric node pairs in the giant component of Erdős-Rényi graphs.

THEOREM 5.2. *For the Erdős-Rényi model, $p = \Omega(\frac{1}{n})$ and $p \leq \frac{1}{2}$, 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 $p \leq \frac{1}{2}$. 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:

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

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

$$\frac{P_{even}}{P_{odd}} = \frac{\sum_{j=1}^{\lfloor \frac{n-1}{2} \rfloor} \frac{\lambda^{2j}}{(2j)!} e^{-\lambda}}{\sum_{j=1}^{\lfloor \frac{n-2}{2} \rfloor} \frac{\lambda^{2j+1}}{(2j+1)!} e^{-\lambda}} \rightarrow \frac{e^\lambda + e^{-\lambda}}{e^\lambda - e^{-\lambda}} \rightarrow 1 \quad (3)$$

Therefore, $P_{even} \rightarrow \frac{1}{2}$ and $P_{odd} \rightarrow \frac{1}{2}$. 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:

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

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 $i \geq 0$, is equal to that of N_b 's. With the same derivation, we can get that $P_{A_2} \geq \sqrt{\frac{2}{\pi n_k}}$. With the same analysis, we can also derive that $P_{A_3} \geq \sqrt{\frac{4}{\pi n_k}}$, $P_{A_4} \geq \sqrt{\frac{8}{\pi n_k}}$ and $P_{A_5} \geq \sqrt{\frac{16}{\pi n_k}}$, 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 $i \geq 0$, is equal to that of N_b 's respectively.

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

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

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:

$$P_{(N_a, N_b) \in \text{sym}} \leq \frac{32}{(\pi n_{k_0})^{\frac{5}{2}}} = o\left(\frac{1}{n^2}\right). \quad (6)$$

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

Otherwise, for any constant $\epsilon > 0$ and two nodes N_1, N_2 in the network, let B be the event: $deg(N_1) = deg(N_2) \geq (\frac{1}{2} + \epsilon)n$. Since $p \leq \frac{1}{2}$, then $\lambda = (n-1)p < \frac{n}{2}$ and we have:

$$\begin{aligned} P_B &= \sum_{i=(\frac{1}{2}+\epsilon)n}^{n-1} \left(\frac{\lambda^i}{i!} e^{-\lambda} \right)^2 \leq \max_{i \geq (\frac{1}{2}+\epsilon)n} \left(\frac{\lambda^i}{i!} e^{-\lambda} \right) \sum_{i=(\frac{1}{2}+\epsilon)n}^{n-1} \frac{\lambda^i}{i!} e^{-\lambda} \\ &< \left(\frac{\lambda}{\frac{1+\epsilon}{2}n} \right)^{\frac{\epsilon}{2}n} e^{-\lambda} < \left(\frac{1}{1+\epsilon} \right)^{\frac{\epsilon}{2}n}. \end{aligned} \quad (7)$$

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 neighbors, 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 $P_f \leq \sum_{d=2}^{(\frac{1}{2}+\epsilon)n} P_{deg=d} \binom{d}{d-n_f} p^{d-n_f}$, 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

$$\begin{aligned}
P_{(N_a, N_b) \in \text{sym}} &< \prod_{k=1}^l \frac{32}{(\pi n_k)^{\frac{5}{2}}} \\
&\leq \frac{32}{\pi^{\frac{5}{2}} (\sum_{k=1}^l n_k - l)^{\frac{5}{2}}} \prod_{k=1}^{l-1} \frac{1}{\sqrt{\pi}} \\
&= \frac{32}{\pi^{\frac{5}{2}} \left(\pi^{\frac{l-1}{5}} (\sum_{k=1}^l n_k - l) \right)^{\frac{5}{2}}} \quad (8) \\
&< \frac{32}{\pi^{\frac{5}{2}} \left(\sum_{k=1}^l n_k \right)^{\frac{5}{2}}} \\
&= o\left(\frac{1}{n^2}\right)
\end{aligned}$$

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

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

□

5.2 Partially Sampled Case

The partially sampled case is relatively difficult to analyze. 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.

REFERENCES

- [1] Béla Bollobás. 1998. *Random Graphs*. Springer New York, New York, NY, 215–252. https://doi.org/10.1007/978-1-4612-0619-4_7
- [2] Ben D. Macarthur, Rubén J. Sánchez-García, and James W. Anderson. 2008. Symmetry in complex networks. *Discrete Applied Mathematics* 156, 18 (2008), 3525–3531.
- [3] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. 2002. Network Motifs: Simple Building Blocks of Complex Networks. *Science* 298, 5594 (2002), 824–827. <https://doi.org/10.1126/science.298.5594.824> arXiv:<https://science.sciencemag.org/content/298/5594/824.full.pdf>
- [4] Ronald C. Read and Robin J. Wilson. 1998. *An atlas of graphs*. Clarendon Press Oxford University Press New York (1998).