2020 FALL AI607 HW5: Final Exam

Fanchen Bu [20194185]

December 9, 2020

1 Linear Threshold Model

Recall the *linear threshold model*, which can be described below:

- Consider a graph G where every node starts with behavior B.
- A small set S of early adopters adopt A and never changes their behavior.
- Each of the other nodes switches from B to A if and only if at least a fraction q of their neighbors have adopted A.
- The process terminates if no node changes their behavior.

Definition 1.1 *A* ***cluster of density*** *p is defined as a set of nodes such that each node in the set has at least a p fraction of its neighbors in the set.*

Definition 1.2 *We say that a* ***complete cascade*** *happens if eventually every node switches from B to A.*

1.1 Questions

1. Prove or disprove the following claim: *"If there exists a cluster of density greater than $1 - q$ such that no early adopter belongs to the cluster, then the set of initial adopters will not cause a complete cascade."*
2. Prove or disprove the following claim: *"If a set of initial adopters does not cause a complete cascade, there must exists a cluster of density greater than $1 - q$ such that no early adopter belongs to the cluster."*

1.2 Answer

1.2.1 Q1

The claim is true. Suppose there exists C, a cluster of density greater than $1 - q$ such that no early adopter belongs to the cluster, we will prove that none of the nodes in it switches from B to A, and thus a complete cascade cannot happen. First, we claim that any node in C cannot switch from B to A if no other nodes in it have adopted A. To see this, as C has density $p > 1 - q$, by definition it means that each node in it has at least a p fraction of its neighbors in the set, in other words, it has at most an $1 - p < q$ fraction of its neighbors outside the set. However, any node that is not an early adopter switches from B to A if and only if at least a fraction q of their neighbors have adopted A, which cannot hold if none of the nodes in C have adopted A, even if all the other nodes outside it have already adopted A, completing the proof of our claim. With this claim, we assume that a complete cascade can happen. Then there exists a step when the first node that adopts A appears in C, however, this cannot happen as before the first node appears, all nodes in C have not adopted A, thus by our claim, none of the nodes in C can switch from B to A, completing the proof.

1.2.2 Q2

The claim is true. Suppose a set of initial adopters does not cause a complete cascade, let M denote the set of nodes that do not change their behavior in the whole process, which should be nonempty by the assumption. Note that all the other nodes in G that are not in M should have adopted A by the end of the process. We claim that M is a cluster of density $p > 1 - q$ and clearly no early adopter belongs to M. To see this, assume the opposite, which means there exists a node $v \in M$ such that v has a $p' < p$ fraction of its neighbors in M. Clearly, we must prove the case for $p' \leq 1 - q$ as p can be arbitrarily close to $1 - q$. However, if $p' \leq 1 - q$, which means that v has an $1 - p' \geq q$ fraction of its neighbors outside M, while we know that all the other nodes outside M have adopted A by the end of the process. Therefore, v must have switched from B to A because at least a fraction q if its neighbors have adopted A. By contradiction, we complete the proof of our claim, and thus complete the whole proof.

2 Triangle Counting

Consider a graph G. G contains τ triangles (i.e., 3-cliques), and k (unordered) pairs of the triangles share an edge. We delete each edge in G independently with probability $1 - p$. Let G' be the remaining graph, and let X be the number of triangles in G'.

2.1 Questions

1. Prove or disprove the following claim: $\mathbb{E}[X] = p^3\tau$.
2. Prove or disprove the following claim: $\mathrm{Var}[X] = \tau(p^3 - p^6) + 2k(p^5 - p^6)$.

2.2 Answer

2.2.1 Q1

The claim is true. Let $\{A_1, \ldots, A_\tau\}$ denote the set of triangles in G. For each $i \in [\tau]$, let X_i be the indicator random variable for A_i remaining in G'. Clearly, $X = \sum_i X_i$. For each A_i, it remains in G' if and only if none of its three edges is deleted, which happens with probability p^3, i.e., $\mathbb{E}[X_i] = p^3$, thus by linearity of expectation, we have

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X_i] = p^3\tau,$$

completing the proof.

2.2.2 Q2

The claim is true. Recall $X = \sum_i X_i$, so we have

$$\mathrm{Var}[X] = \sum_i \mathrm{Var}[X_i] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j],$$

where the second sum is over ordered pairs and the covariance is defined by

$$\mathrm{Cov}[Y, Z] = \mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z].$$

First, for each X_i, $\mathrm{Var}[X_i] = p^3(1-p^3) = p^3 - p^6$ as it's simply binomial. In terms of covariances, if X_i and X_j are independent, i.e., A_i and A_j share no edges, then $\mathrm{Cov}[X_i, X_j] = 0$; otherwise we have

$$\mathrm{Cov}[X_i, X_j] = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] = p^5 - p^6,$$

where p^5 comes from the fact that we should keep 5 edges to make two triangles that share an common edge remain. Now we can conclude that

$$\mathrm{Var}[X] = \tau(p^3 - p^6) + 2k(p^5 - p^6),$$

where $2k$ is from the sum over ordered pairs which counts each unordered pair twice, completing the proof.

3 Kronecker Model

Consider an initiator graph be G_1 without self-loops. In the (deterministic) Kronecker model, by recursion, we produce successively large graphs $G_2, G_3, \cdots$.

3.1 Questions

1. Prove or disprove the following claim: *"If we let $d_k^{\max}$ be the maximum degree of nodes in G_k, then $d_k^{\max} = (d_1^{\max})^k$."*

2. Prove or disprove the following claim: *"If we let D_k be the diameter of G_k, then $D_k = (D_1)^k$."*

3.2 Answer

3.2.1 Q1

The claim is true. First, we prove the following lemma.

Lemma 3.1 *For any simple graph G on n vertices with corresponding degree sequence $d_1, \dots, d_n$ and another simple graph H on m vertices with corresponding degree sequence $f_1, \dots, f_m$, $G \otimes H$ is a graph on mn vertices with degree sequence $g_{11}, \dots, g_{mn}$, where $g_{ij} = f_i d_j$ for each $i \in [n]$ and $j \in [m]$. Specifically, let $d^* = \max_{i \in [n]} d_i$ and $f^* = \max_{i \in [m]} f_i$ denote the maximum degree of G and H, respectively, we have the maximum degree of $G \otimes H$, $g^* = \max_{i,j} g_{ij} = d^* f^*$.*

Proof. Consider the adjacent matrix $A \in \{0,1\}^{n\times n}$ of G and the adjacent matrix $B \in \{0,1\}^{m\times m}$ of H, by the definition of Kronecker product, the adjacent matrix C of $G \otimes H$ has size $mn \times mn$, where

$$C_{i,j} = A_{\lceil i/m\rceil,\lceil j/m\rceil} B_{(i-1)\%m+1,(j-1)\%m+1}.$$

As both G and H are simple graph, by symmetry we fix i, and compute its corresponding degree

$$\sum_j C_{i,j} = \sum_j A_{\lceil i/m\rceil,\lceil j/m\rceil} B_{(i-1)\%m+1,(j-1)\%m+1},$$

where each term in the summation is 1 if and only if both $A_{\lceil i/m\rceil,\lceil j/m\rceil}$ and $B_{(i-1)\%m+1,(j-1)\%m+1}$ are 1. As j enumerates over $[mn]$, $(\lceil j/m\rceil, (j-1)\%m+1)$ enumerates over $[n]\times[m]$, which gives

$$\sum_j C_{i,j} = d_{\lceil i/m\rceil} f_{(i-1)\%m+1}$$

as degree is equal to the number of 1 in the corresponding column or row in the adjacent matrix. Besides, if we let i enumerate over $[mn]$, then $(\lceil i/m\rceil, (i-1)\%m+1)$ also enumerates over $[n]\times[m]$, completing the proof. ■

As $G_k = G_{k-1} \otimes G_1$ for each $k > 1$, by Lemma 3.1 we have

$$\begin{aligned} d_k^{\max} &= d_{k-1}^{\max} d_1^{\max} \\ &= d_{k-2}^{\max}(d_1^{\max})^2 \\ &= \dots \\ &= (d_1^{\max})^k, \end{aligned}$$

completing the proof.

3.2.2 Q2

The claim if false. Consider this simple case where G_1 is a graph on vertex set $[3]$ with two edges 12 and 23. Clearly, G_1 has diameter 2 and the adjacent matrix of G_1 is

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

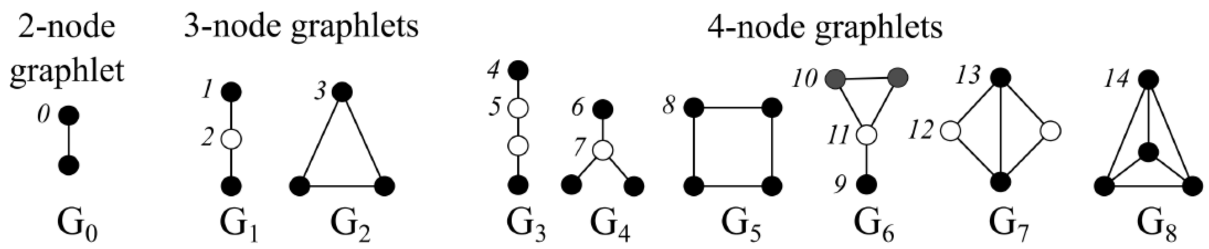
However, the adjacent matrix of $G_2 = G_1 \otimes G_1$ is

$$\begin{pmatrix} 0&0&0&0&1&0&0&0&0 \\ 0&0&0&1&0&1&0&0&0 \\ 0&0&0&0&1&0&0&0&0 \\ 0&1&0&0&0&0&0&1&0 \\ 1&0&1&0&0&0&1&0&1 \\ 0&1&0&0&0&0&0&1&0 \\ 0&0&0&0&1&0&0&0&0 \\ 0&0&0&1&0&1&0&0&0 \\ 0&0&0&0&1&0&0&0&0 \end{pmatrix}$$

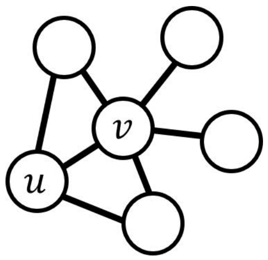
and we can easily check that G_2 is not even connected (as even nodes are only connected with even nodes and likewise the odd nodes are only connected with odd ones), i.e., the diameter of G_2 is ∞, implying that the claim is false.

4 Graphlet

Below, we show all 2-, 3-, and 4-node graphlets, from which we can calculate a Graphlet Degree Vector of size 14 (GDV-14).



Consider the following graph G:

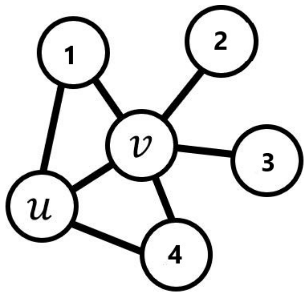


4.1 Questions

1. What is the GDV-14 of node u?
2. What is the GDV-14 of node v?

4.2 Answer

First, we label the other nodes as follows:



4.2.1 Q1

For G_0, u acts in orbit position 0 in the induced subgraph(s) of G on $\{u, 1\}, \{u, v\}, \{u, 4\}$.
For G_1, u acts in orbit position 1 in the induced subgraph(s) of G on $\{u, v, 2\}, \{u, v, 3\}$ and acts in orbit position 2 in the induced subgraph(s) of G on $\{u, 1, 4\}$.
For G_2, u acts in orbit position 3 in the induced subgraph(s) of G on $\{u, v, 1\}, \{u, v, 4\}$.

For G_3, there is no induced subgraph of G containing u that is isomorphic to G_3.
For G_4, u acts in orbit position 6 in the induced subgraph(s) of G on $\{u, v, 2, 3\}$.
For G_5, there is no induced subgraph of G containing u that is isomorphic to G_5.
For G_6, u acts in orbit position 10 in the induced subgraph(s) of G on $\{u, v, 1, 2\}, \{u, v, 1, 3\}, \{u, v, 2, 4\}, \{u, v, 3, 4\}$.
For G_7, u acts in orbit position 13 in the induced subgraph(s) of G on $\{u, v, 1, 4\}$.
For G_8, there is no induced subgraph of G containing u that is isomorphic to G_8.
Thus, the final GDV table is:

Orbit	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
GDV(u)	3	2	1	2	0	0	1	0	0	0	4	0	0	1	0

4.2.2 Q2

For G_0, v acts in orbit position 0 in the induced subgraph(s) of G on $\{u, v\}, \{v, 1\}, \{v, 2\}, \{v, 3\}, \{v, 4\}$.
For G_1, v acts in orbit position 2 in the induced subgraph(s) of G on $\{u, v, 2\}, \{u, v, 3\}, \{v, 1, 2\}, \{v, 1, 3\}, \{v, 1, 4\}, \{v, 2, 3\}, \{v, 2, 4\}, \{v, 3, 4\}$.
For G_2, v acts in orbit position 3 in the induced subgraph(s) of G on $\{u, v, 1\}, \{u, v, 4\}$.
For G_3, there is no induced subgraph of G containing v that is isomorphic to G_3.
For G_4, v acts in orbit position 7 in the induced subgraph(s) of G on $\{u, v, 2, 3\}, \{v, 1, 2, 3\}, \{v, 1, 2, 4\}, \{v, 1, 3, 4\}, \{v, 2, 3, 4\}$.
For G_5, there is no induced subgraph of G containing v that is isomorphic to G_5.
For G_6, v acts in orbit position 11 in the induced subgraph(s) of G on $\{u, v, 1, 2\}, \{u, v, 1, 3\}, \{u, v, 2, 4\}, \{u, v, 3, 4\}$.
For G_7, v acts in orbit position 13 in the induced subgraph(s) of G on $\{u, v, 1, 4\}$.
For G_8, there is no induced subgraph of G containing v that is isomorphic to G_8.
Thus, the final GDV table is:

Orbit	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
GDV(v)	5	0	8	2	0	0	0	5	0	0	0	4	0	1	0