CSC303 Winter 2020

Sample solution to Assignment #1

Question 1

(a) Let $A(G)$ be the adjacency matrix of our graph (so $a_{ij} = 1$ iff we have a connection from i to j). Let us define $B(G) := A(G) + I$ and $C := B(G)^2$. Let us use $\mathbb{I}_{x=k}$ to denote the indicator function ($\mathbb{I}_{x=k}$ is 1 when $x = k$ and is 0 otherwise). Then by the definition of matrix multiplication:

$$c_{ij} = [A(G) + I]_{i:}[A(G) + I]_{:j}$$

$$= \sum_{k=1}^{n}(a_{ik} + \mathbb{I}_{i=k})(a_{kj} + \mathbb{I}_{j=k})$$

$$= \sum_{k \neq i,j} a_{ik}a_{kj} + \begin{cases} (a_{ii}+1)(a_{jj}+1), & i = j \\ (a_{ii}+1)a_{ij} + a_{ij}(a_{jj}+1), & i \neq j \end{cases}$$

Note that as our initial graph does not have self loops, $a_{ii} = a_{jj} = 0$. Therefore:

$$c_{ij} = \sum_{k \neq i,j} a_{ik}a_{kj} + \begin{cases} 1, & i = j \\ 2a_{ij}, & i \neq j \end{cases}$$

Now, observe that c_{ij} is the number of length 2 paths from i to j (as calculated by $\sum_{k \neq i,j} a_{ik}a_{kj}$) plus either 1 (if $i = j$) or plus 2 (if $i \neq j$ and a path from i to j exists).
Note that this is the number of length 2 paths from i to j on the graph represented by $B(G)$ (i.e. our original graph, plus self-loops - note that if $i \neq j$ and there is a path between them in G, then we have 2 paths once we introduce self loops, namely i, i, j and i, j, j).
Alternatively, if we consider the graph G then we can see that c_{ij} is the number of length 0 paths from i to j plus twice the number of length 1 paths from i to j, plus the number of length 2 paths.

(b) The (i, j) entry of $B(G)^k$ is non-zero if and only if there's a path of length which is smaller than or equal to k between vertex i and vertex j. Since the maximum path length in the graph is d, the smallest value of k is d.

Question 2

(a) Two days. First, D and G don't have common friends so they cannot be friends by triadic closure in one day. Second, D can connect with F on the first day and then connect with G on the second day by triadic closure.

(b) There are two ways of connecting D and K in two days, either through F or B. Then we calculate the probability of each triadic closure. For simplicity, we use $\Pr[e_{AB}]$ represents the probability of forming an edge between A and B.
First day:

Pr[e_{KF}] = 1/2, through G.
Pr[e_{KB}] = 1/2, through H.
Pr[e_{DB}] = 1/2, through A.
Pr[e_{DF}] = 1/2 + 1/2 - 1/2 * 1/2 = 3/4, through C or A and apply addition rule of probability.
Second day:
The probability of connecting D and K through F is equal to Pr[e_{DF}] * Pr[e_{KF}] * 1/2= 3/16.
The probability of connecting D and K through B is equal to Pr[e_{DB}] * Pr[e_{KB}] * 1/2= 1/8.
Hence, Pr[e_{DK}] = 1/8 + 3/16 - 1/8 * 3/16 = 37/128.

Question 3

(a) Strong: $(D, C), (A, B), (E, F)$
Weak: $(A, C), (A, D), (B, G), (B, F), (A, E)$

(b) Since $(B, E), (A, F), (A, G), (B, C), (B, D) \notin E$, $(B, G), (A, E), (B, F), (A, C), (A, D)$ must be weak edges by strong triadic closure property. Then we can label $(F, E), (D, C)$ as strong edges

(c) Strong: $(G, B), (F, E), (A, D), (A, C), (D, C)$
Weak: $(B, F), (B, A), (A, E)$

Question 4

(a) There's no bridge in the graph since the graph is still connected after removing one edge. $(3, 10)$ is a local bridge of span 3.

(b) Note: For path from 1 to 7, we have $1-3-5-7$, $1-2-5-7$ and $1-3-10-7$. By symmetry, there are three shortest paths from 11 to 5.

First node in pair	only 1 via $(5,7)$	only 1 via $(3,10)$	1 via $(5,7)$ & 1 via $(3,10)$	2 via $(5,7)$ & 1 via $(3,10)$
1		$\{8,9,10,11\}$		$\{7\}$
2	$\{7\}$	$\{10,11\}$	$\{8,9\}$	
3		$\{8,9,10,11\}$	$\{7\}$	
4	$\{7\}$	$\{10,11\}$	$\{8,9\}$	
5	$\{7,8,9\}$		$\{10\}$	$\{11\}$
total flow	5	12	6	2

Total $(5,7)$ flow: $5 + 0.5 \times 6 + 2 \times \frac{2}{3} = \frac{28}{3}$
Total $(3,10)$ flow: $12 + 0.5 \times 6 + 2 \times \frac{1}{3} = \frac{47}{3}$

Question 5

(a) There's strong evidence of homophily. In the graph, there are 50% volleyball nodes and 50% hockey nodes. Consider any edge (u, v) in the graph. If the engagement in the sport has no effect on friendship, then the probability of u and v are engaged in different sports is 2 * 1/2 * 1/2 = 1/2. However, in the graph, there are only 3 edges, out of 24 edges, which are between teenagers engaged in different sports. The probability 1/8 is significantly less than 1/2.

(b) The network of relationships has formed based on sports.

(c) (U, V)

(d) Only teenagers who have at least one friend engaged in a different sport can decide to engage a new sport. As a result, all teenagers on the boundary, B, D, F, X, Y, Z, have the same probability 1/4 to engage in a new sport.

(e) It will spread to all hockey players. First, X, Y will adopt τ since they are friends with both W and U. Then, Z will adopt τ due to Y and W. Finally, V will adopt τ due to Z and W. However, it will not spread to volleyball players since all of them has at most one hockey player friend.

Question 6

(a)

- Triadic closure
  - A and F, through friend E and interest d: $P = (1 - 0.5^1)\frac{1+1}{3} = \frac{1}{3}$
  - B and C, through friend H and interest e: $P = (1 - 0.5^1)\frac{1+1}{3} = \frac{1}{3}$
  - B and E, through friend F and interest d: $P = (1 - 0.5^1)\frac{1+1}{3} = \frac{1}{3}$
  - B and G, through friend F, H and interest e: $P = (1 - 0.5^2)\frac{1+1}{3} = 0.5$
- Focal closure
  - C, F, through Club 2 and interest e: $P = (1 - 0.8^1) = 0.2$
- Membership closure
  - A and Club 1, through D and E: $P = 1 - 0.7^2 = 0.51$
  - B and Club 2, through F: $P = 1 - 0.7 = 0.3$
- Triadic Focal closure
  - C, G,
    through Club 2 and interest e: $p_f = 1 - 0.8 = 0.2$
    through friend H and interest e: $p_f = 1 - 0.8 = 0.2$, $p_t = (1 - 0.5^1)\frac{1+1}{3} = \frac{1}{3}$.
    So, $P = 1 - (1 - 0.2)(1 - 1/3) = 7/15$

(b) From the probability from (a), A is most likely to join Club 1 since A has two friends at Club 1.
Club 1 is popular among students who share interest in dentistry while Club 2 is popular among students who share interest in engineering.

(c) (D, E) is most embedded since they have two common neighbours. ($(A, D), (D, E), (A, E)$ is also accepted.)

(d) There's strong evidence of homophily. In the graph, there are 37.5% nodes with interest in d and 37.5% nodes with interest in e. Consider any edge (u, v) in the graph. If the career interest has no effect on friendship, then the probability of u and v have different career interest is 2 * 0.375 * 0.375 = 0.28125. However, in the graph, there are no edges between students with different interest. There are only one edge $de - d$ edge and one $de - de$ edge and two $de - e$ edges, but this cannot be used to justify the homophily since two endpoints of such an edge share common interest.

Question 7

Conclusions:

- The final ”% similar” is much higher than ”% similar-wanted”. One possible explanation is that: one person leaves the grid may cause its neighbour falling its threshold, so it’s not stable to have ”% similar” close to the threshold.

- Based on the results, the number of ticks increases (it takes more time to converge) as the ”% similar wanted” increases. One explanation: The gap is larger so it’s less likely to be satisfied, which means more time to converge. Also, it might not be enough to change some person’s position (locally) to increase the ”% similar” (a global property). Instead, we need to change the global pattern, which takes more time.

- The number of ticks increases more rapidly when $N = 2500$ than $N = 900$. The explanation is similar to the previous one. Increasing ”% similar” requires the global change of the pattern. Intuitively, the global change takes more time when we have a larger sample size. Also, it takes more time to make more people satisfied with the same threshold.

- With the same ”% similar-wanted”, the final ”% similar” is higher when $N = 900$ than $N = 2500$. One possible explanation: A smaller sample size means more flexibility which allows more possible moves of a person, so it results in a higher ”% similar”.

	$N = 900$		$N = 2500$	
	%-Sim	Ticks	%-Sim	Ticks
$t = 20\%$	Avg. 65.44 Min. 64.7 Max. 66.5	Avg. 6.2 Min. 6 Max. 7	Avg. 56.06 Min. 54.1 Max. 57.8	Avg. 8.4 Min. 6 Max. 12
$t = 30\%$	Avg. 79 Min. 78 Max. 79.8	Avg. 10.4 Min. 9 Max. 12	Avg. 75.3 Min. 73.6 Max. 76.4	Avg. 17 Min. 11 Max. 22
$t = 50\%$	Avg. 91.25 Min. 89.3 Max. 92.4	Avg. 14.7 Min. 7 Max. 19	Avg. 86.34 Min. 85.5 Max. 87.1	Avg. 30.1 Min. 15 Max. 44
$t = 70\%$	Avg. 99.8 Min. 99.7 Max. 100	Avg. 51.8 Min. 44 Max. 60	Avg. Min. Max.	Avg. Min. Max.