

Solutions to Homework II

Problem 1: This problem has two parts.

- (a) Consider a class of elementary school students consisting of 9 boys and 12 girls. Suppose a social network on this group exhibits extreme gender homophily; that is, it has no cross-gender edges. Compute the *maximum* number of possible edges in the social network.
- (b) Consider a set of high school students consisting of 120 girls and 80 boys. A social network on this set has a total of 1000 edges. Suppose the number of cross-gender edges in this network is exactly 40% of the value predicted by the random mixing model discussed in class. Find the number of cross-gender edges in the network.

Solution:

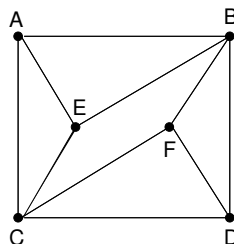
Part (a): When the network exhibits extreme gender homophily, each edge is between a pair of boys or a pair of girls. Recall that for a graph with n nodes, the maximum number of possible edges $= n(n - 1)/2$. (This happens when the graph is an n -clique.)

Thus, the maximum number of possible edges among boys $= 9(9 - 1)/2 = 36$ and the maximum number of possible edges among girls $= 12(12 - 1)/2 = 66$. Therefore, the maximum number of possible edges in the network is $36 + 66 = 102$.

Part (b): Since the social network has 120 girls and 80 boys, the probability p of choosing a girl is $120/(120+80) = 0.6$ and the probability q of choosing a boy is $1.0 - 0.6 = 0.4$. Under the random mixing model, the expected number of cross-gender edges is $2pq|E|$, where E is the total number of edges. Here, since $|E| = 1000$, the expected number of cross-gender edges is $2 \times 0.6 \times 0.4 \times 1000 = 480$. However, the number of cross-gender edges in the network is 40% of the expected value. So, the actual number of cross-gender edges in the network is $480 \times 0.4 = 192$.

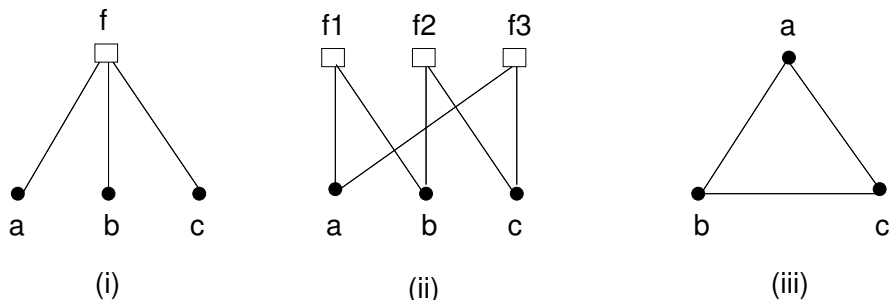
Problem 2: Recall that an **affiliation network** is a bipartite graph with two sets of nodes: one set P represents people and the other set F represents focal points. Further, each edge is between a node in P and one in F . Also recall that given an affiliation network G_A , one can construct a **projected network** G_P of G_A as follows: the set of nodes for G_P is P itself and G_P has an edge between a pair of nodes x and y in P if and only if x and y have at least one common focal point in F . This problem has two parts.

- (a) Show two different affiliation networks G_A^1 and G_A^2 such that the projected networks for the two are *identical* (i.e., the two projected networks have the same set of edges).
- (b) Consider the following social network G . Construct an affiliation network G_A such that G is the projected network of G_A . The network G_A must use *at most* 4 focal points.

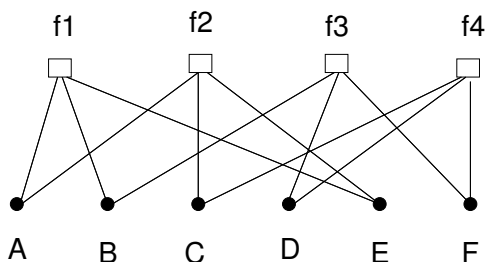


Solution:

Part (a): Figures (i) and (ii) below represent two different affiliation networks G_A^1 and G_A^2 . (In these figures, nodes corresponding to people are shown as dark circles and those corresponding to focal points are shown as hollow rectangles.) These two affiliation networks have the same projected network G given by (iii).



Part (b): The following figure shows one affiliation network whose projected network is shown in the statement of this problem.



Problem 3: For any positive integer n , prove that there is an undirected graph with $N \geq n$ nodes and $\Omega(N^2)$ edges such that the clustering coefficient of each node is zero. (An undirected graph G with N nodes is **dense** if the number of edges in G is $\Omega(N^2)$. This problem points out that dense graphs may have small clustering coefficients.)

Solution: We will show how to construct such a graph. If n is even, let $N = \max\{n, 4\}$; otherwise, let $N = \max\{4, n + 1\}$. Note that N is even and $\geq n$. Note also that $N \geq 4$. (This allows us to avoid having nodes of degree 0 or 1 in the graph; recall that for such nodes, the value of clustering coefficient is *undefined*.)

Let $N = 2k$. Construct the complete bipartite graph with k nodes on each side of the bipartition. The number of edges in this graph is $k^2 = (N/2)^2 = N^2/4 = \Omega(N^2)$.

For any node x in the graph, the $k = N/2$ neighbors of x are on the opposite side. Since $N \geq 4$, $k \geq 2$; thus, there is no node with degree 0 or 1 in the graph. Moreover, for any node x , the neighbors of x form an independent set (i.e., there is no edge between any pair of neighbors of x). In other words, the clustering coefficient of each node is 0.

(over)

Problem 4: Let n be an even positive integer. Suppose G is an undirected graph with n nodes such that each node of G has a clustering coefficient of zero. Prove that the number of edges in G is at most $n^2/4$.

Solution: We will use the following two well known facts in the proof.

Fact 1: Let G be an undirected graph such that the clustering coefficient of each node is 0. Then, the neighbors of each node v of G form an independent set (i.e., there is no edge between any pair of neighbors of v).

Fact 2: For positive numbers p and q , $\sqrt{pq} \leq (p + q)/2$. (That is, for any pair of positive numbers, their geometric mean is at most their arithmetic mean.)

Proof of the required result: Let G be a graph with n nodes in which the clustering coefficient of each node is zero. We must show that the number of edges in G is at most $n^2/4$.

Let v be a node of *maximum* degree in G , and let d denote the degree of v . Let A denote the neighbors of v (not including v). By Fact 1, there are no edges between any pair of nodes in A . Thus, each of the edges of G is incident on some node in $V - A$. The number of nodes in $V - A$ is $n - d$ and each of these nodes has degree at most d . Therefore, the number of edges in G is at most $d(n - d)$. Now,

$$\begin{aligned} d(n - d) &= \left[\sqrt{d(n - d)} \right]^2 \\ &\leq \left[(d + n - d)/2 \right]^2 \quad (\text{from Fact 2}) \\ &= n^2/4. \end{aligned}$$

Thus, the number of edges in G is at most $n^2/4$, as required. ■

Note: This result is a special case of Turán's theorem in Graph Theory.
