Assignment 2

Start Assignment

Due Monday by 11:59pm **Points** 100 **Submitting** a file upload **File Types** zip
Available Mar 1 at 12am - Mar 21 at 11:59pm 21 days

1. (30 points) Given an undirected and weighted graph $G =< V, E >$, where V is the set of nodes, $E \subseteq V \times V$ is the edge set, and each edge $(u, v) \in E$ is assigned with a weight w_{uv}. For a subset of nodes $S \subseteq V$, denote by $E(S) = \{(u, v) \mid u \in S, v \in S, (u, v) \in E\}$ the set of edges between the nodes in S. We define the **weighted density** of a subgraph (a subset of nodes) $S \subseteq V$ as

$$\rho(S) = \frac{\sum_{(u,v)\in E(S)} w_{uv}}{|S|}$$

(Note that since (u, v) and (v, u) represent the same edge in graph G, in the above summation each edge (u, v) is only counted once, which means if (u, v) is counted then (v, u) is not counted).

For two subsets $S_1, S_2 \subseteq V$, we define their **weighted cut** as

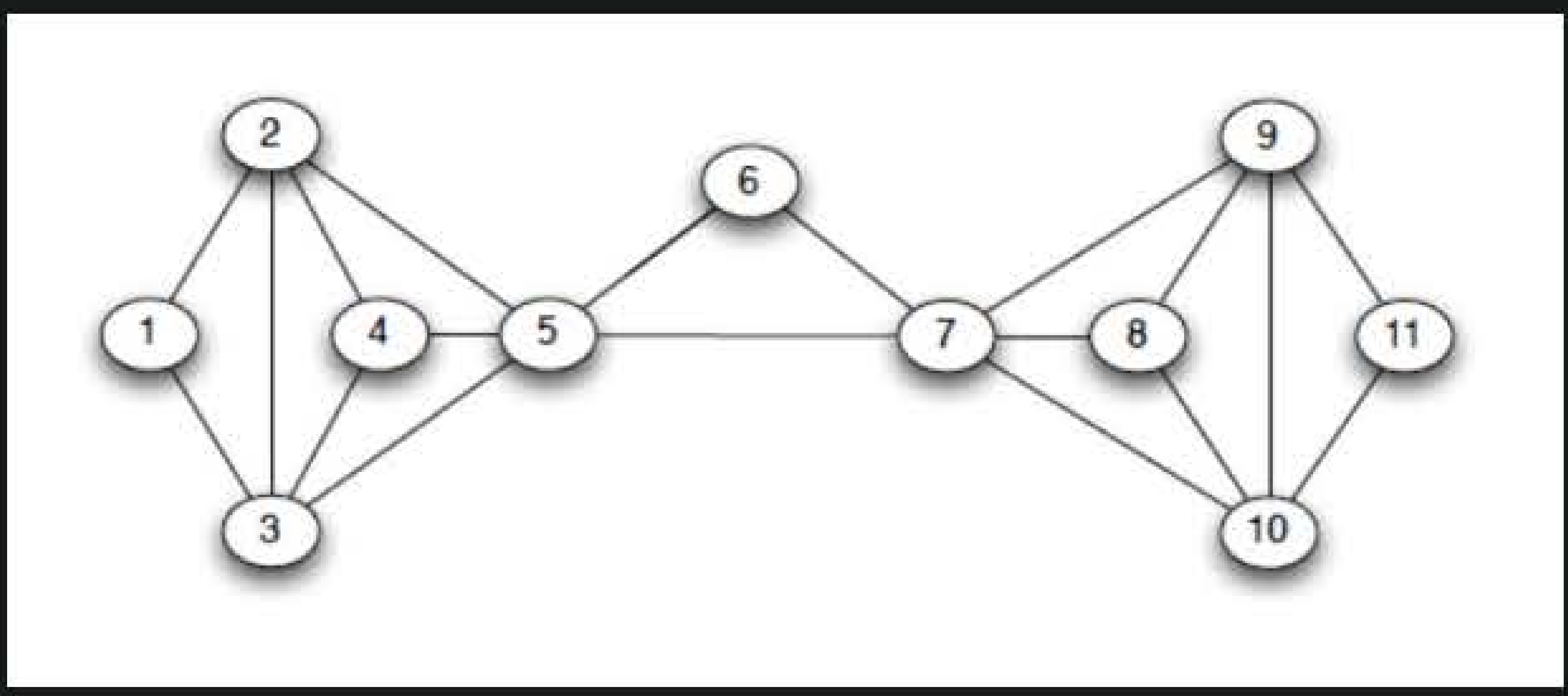
$$WC(S_1, S_2) = \sum_{(u,v)\in E,\ u\in S_1, v\in S_2} w_{uv}$$

A subgraph (subset of nodes) $S \subseteq V$ is a **clique** if for any nodes pair $u, v \in S$ $(u \neq v)$, there is an edge $(u, v) \in E$. A **maximum clique** of G is a clique with the biggest number of nodes among all cliques in G. The **maximum clique size** is the number of nodes in a maximum clique of G.

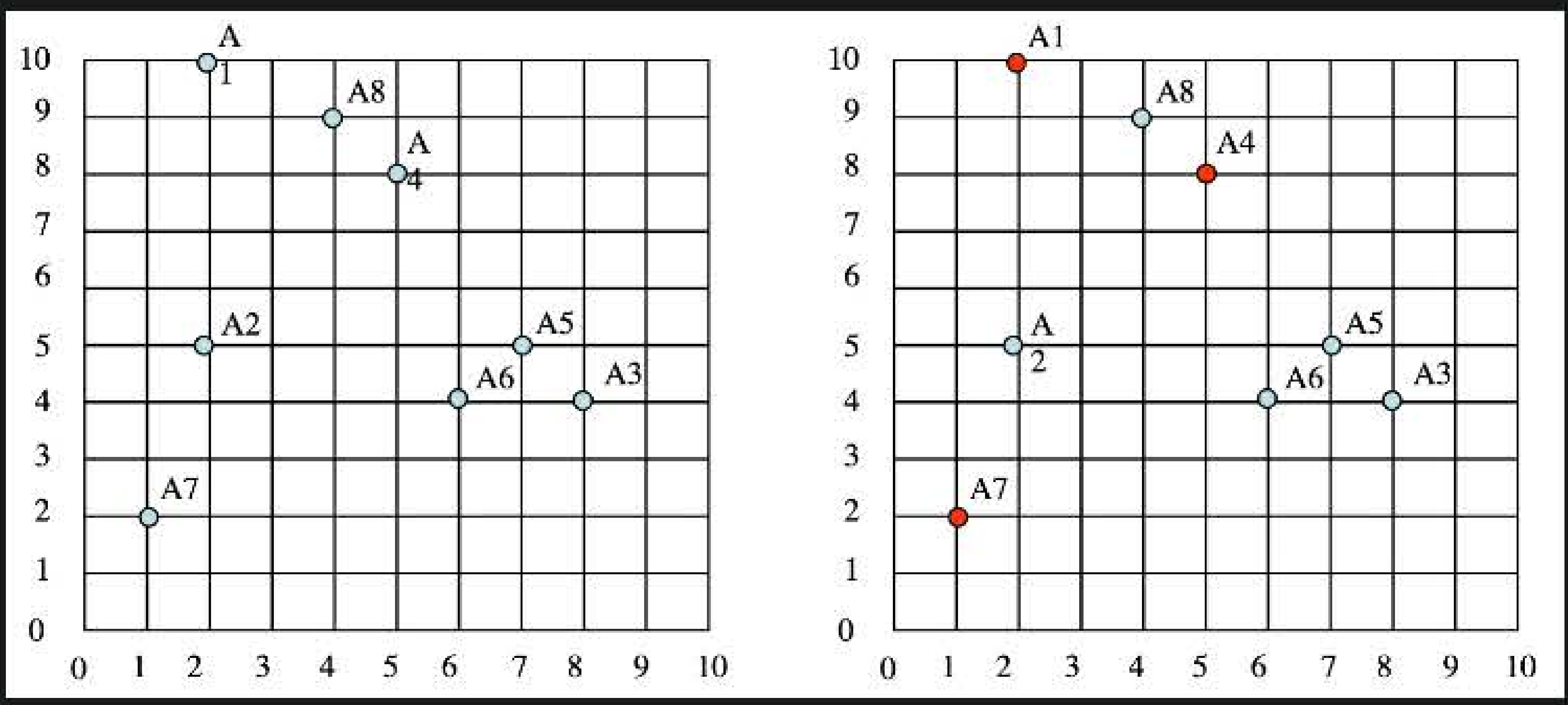
(1) (20 points) Suppose all edges carry positive weights, that is, $\forall (u, v) \in E, w_{uv} > 0$. Please describe how to construct a weighted graph $G' =< V \cup \{s, t\}, E' >$ as well as a positive number γ, such that there exists a subgrph $S \subseteq V$ in G where $\rho(S) \geq \lambda$ *is equivalent to* there is an (s, t)-cut in G', denoted by $\{s\} \cup S_1$ and $(V - S_1) \cup \{t\}$, where $WC(\{s\} \cup S_1, (V - S_1) \cup \{t\}) \leq \gamma$.

(2) (10 points) Let's consider a more general case of the weighted graph G, where some edges in G carry negative edge weights, that is, $\exists (u, v) \in E,\ w_{uv} < 0$. We want to show that finding the densest subgraph w.r.t. the weighted density in a graph G with negative weighted edges can help us identify the maximum clique size of G. Suppose the maximum clique size of G is c. Please figure out how to construct a weighted graph $G' =< V, E' >$ (note that the node set is still V) where for any edge $(u, v) \in E'$, $w_{uv} \in [-|E| - 1, |E| + 1]$, and the weighted density of the densest subgraph in G' (w.r.t. the weighted density) is $\frac{c-1}{2}$.

2. (15 points) For an undirected graph $G = \langle V, E \rangle$, denote by $\mathbf{A}$ its **adjacency matrix** where $A_{ij} = 1$ if $(i, j) \in E$ and $A_{ij} = 0$ otherwise. Let d_i be the degree of node i, that is, $d_i = |\{(i, j) \mid (i, j) \in E\}|$. Given the following graph, build its **modularity matrix** $\mathbf{B}$ where $B_{ij} = A_{ij} - \frac{d_i d_j}{2|E|}$.



3. (15 points) Use the K-Means algorithm and Euclidean distance to cluster the following 8 data points into 3 clusters:
$A1 = (2, 10), A2 = (2, 5), A3 = (8, 4), A4 = (5, 8), A5 = (7, 5), A6 = (6, 4), A7 = (1, 2), A8 = (4, 9)$. Suppose the initial centroids are $A1$, $A4$ and $A7$. The following figure shows the data points. Run the K-Means algorithm until converge. Show the intermediate results (including the objective value, the cluster assignments, and the centroids) of all K-Means iterations, where an iteration consists of re-assigning clusters to data points and then re-calculating centroids.



4. (40 points) Given an undirected graph $G = \langle V, E \rangle$, the **core number** of a node $u \in V$ is the largest integer k such that there exists a subgraph (subset of nodes) $S \subseteq V$ where $u \in S$ and every node in S is connected to at least k other nodes in S.

(1) (15 points) Calculate the core number of each node in the graph in the figure of Question 2.

(2) (10 points) Figure out an algorithm that calculates the core number of each node for any input graph $G = \langle V, E \rangle$. Analyze the time complexity of your algorithm. (**Hint**: recall the Greedy algorithm for computing an approximate densest subgraph)

(3) (15 points)

i). (10 points) Download the file "DBLP.txt" where the first line describes how many nodes and edges are in the DBLP graph, and each line starting from the second line is in the form "$u \quad v$" which means there is an edge between the node u and the node v. Write a program to calculate the core number of each node for the DBLP graph. Output your results in a txt file where every line is in the form "$u : c_u$" indicating that the core number of node u is c_u. Note that similar to Assignment 1, you also need to upload your code and a readme file describing how to run your code.

ii) (5 points) Try to make your program finish running within 10 seconds (including reading the data file, calculating core numbers, and outputting the results to a txt file).