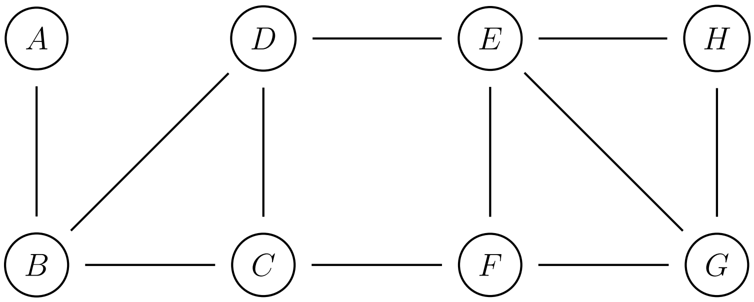
Exercise Sheet 13: Community Detection
Maximilian Marx, Markus Krötzsch
Knowledge Graphs, 2019-01-29, Winter Term 2018/2019

Exercise 13.1. Let $G = \langle V, E \rangle$ be an undirected graph. A *cut* $C = \langle A, B \rangle$ of G is a partition $A \cup B = V$ of V such that there is an edge in G from a vertex $a \in A$ to a vertex $b \in B$. The *cut-set* of C is $S \coloneqq \{\, \{a, b\} \in E \mid a \in A, b \in B \,\}$. The *normalised cut value* of a cut $C = \langle A, B \rangle$ is

$$\frac{\left|\{\, \{a, b\} \in E \mid a \in A, b \in B \,\}\right|}{\left|\{\, \{a, v\} \in E \mid a \in A, v \in V \,\}\right|} + \frac{\left|\{\, \{a, b\} \in E \mid a \in A, b \in B \,\}\right|}{\left|\{\, \{b, v\} \in E \mid b \in B, v \in V \,\}\right|}, \text{ i.e.,}$$

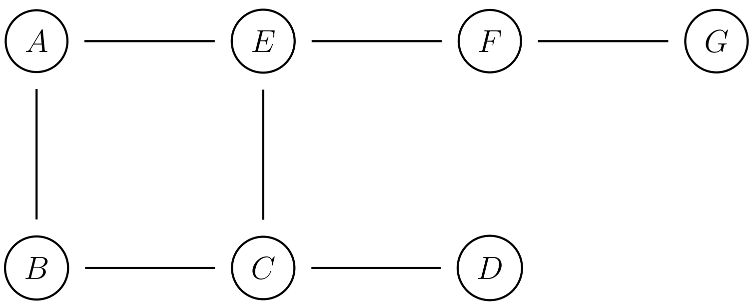
the number of endpoints of edges in the cut-set divided by the number of vertices that have an edge with an endpoint in A or B.

Find the minimal cuts (i) with respect to the cardinality of the cut-set, and (ii) with respect to the normalised cut value of the following graph. Which of these cuts best describes the community structure of the graph?



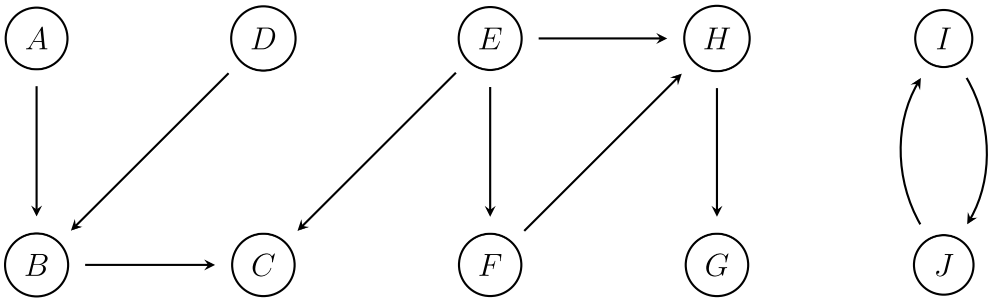
Exercise 13.2. Let $G = \langle V, E \rangle$ be an undirected graph. A *bi-clique* in G consists of two disjoint, nonempty sets $A, B \subseteq V$ such that the induced subgraph of $A \cup B$ in G is a complete bipartite graph, i.e., $\{\, \{a, b\} \mid a \in A, b \in B \,\} \subseteq E$, and no two vertices in A (and B, respectively) are adjacent.

Find all bi-cliques in the following graph. Which bi-cliques are maximal?



Exercise 13.3. Let $G = \langle V, E\rangle$ be a directed graph. For two vertices $v, w \in V$, the *distance* $d(v, w)$ is the length of the shortest directed path from v to w (or ∞ if there is no such path). For a set $S \subseteq V$ of vertices, the *reachable set* is $R(S) := \{ v \in V \mid \exists w \in S.\, d(w, v) < \infty \}$. A *point base* of G is a minimal set $B \subseteq V$ such that $R(B) = V$.

Find a point base for the following graph. How does the point base change when adding the edge $\langle B, E\rangle$?



Exercise 13.4. Write a program that takes as input

- a directed graph in METIS format
- and a dictionary file in the format of Exercise 11.4 mapping vertex IDs to labels,

and uses the Girvan-Newman algorithm to print out all communities on level k of the hierarchical clustering of the input graph.

Modify the program from Exercise 11.4 to extract the subgraph from Wikidata that consists of all edges of the form $t \to s$, where t has a P375 ("space launch vehicle")-statement with value s, and also a P619 ("time of spacecraft launch")-statement with some value. Use your program to print out the communities on level 5 of the hierarchical clustering of this subgraph.

Hint: NetworkX[1] provides an implementation of the GN algorithm.[2]

[1] https://networkx.github.io/

[2] https://networkx.github.io/documentation/latest/reference/algorithms/generated/networkx.algorithms.community.centrality.girvan_newman.html