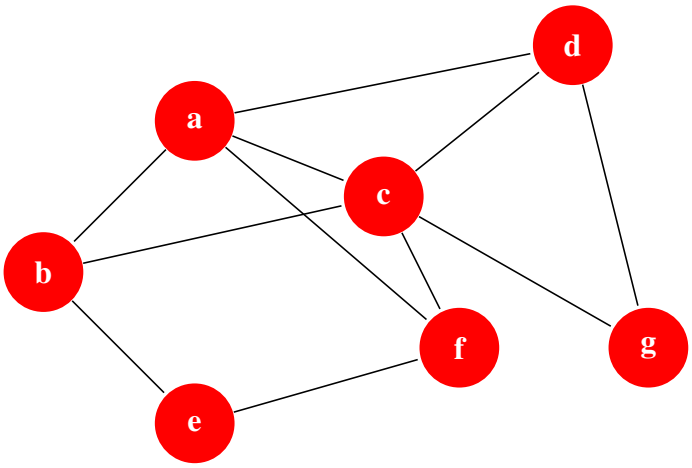
University of Augsburg, Institute of Computer Science
Prof. Dr. P. Fischer
J. Kastner, L. Rudenko

SS 2019
12. July 2019
Sheet 11

Analyzing Massive Data Sets

Exercise 1: Counting Triangles (homework)

The following graph G is given:

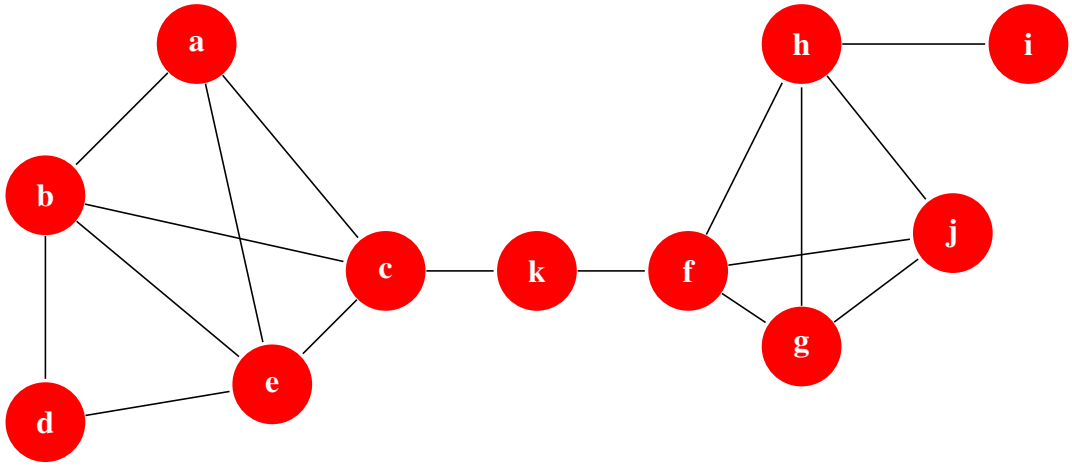


a) Compute the number of triangles in the graph G using the node-centric algorithm for the lowest degree (as described on the slide 18, chapter 8). Provide the execution steps/considerations for all nodes.

b) Write Python code for counting triangles.

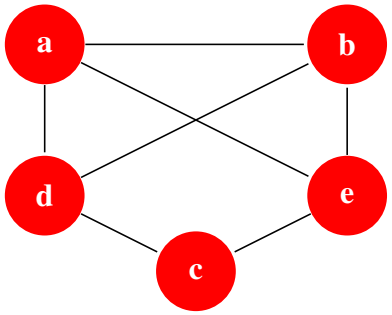
Exercise 2: k-core (homework)

The following graph G is given:



Compute 3-core and belonging component(s) for the given graph G using the algorithm of Batagelj and Zaversnik. You do not need to express the degree computation and bin sorting, results for this stage (i.e., the nodes sorted by degree into bins) are sufficient. Output the calculated 3-core and belonging component(s). Note, than k-core is a **maximal subgraph where** $deg(n) >= k$.

Exercise 3: Girvan-Newman Algorithm (homework)

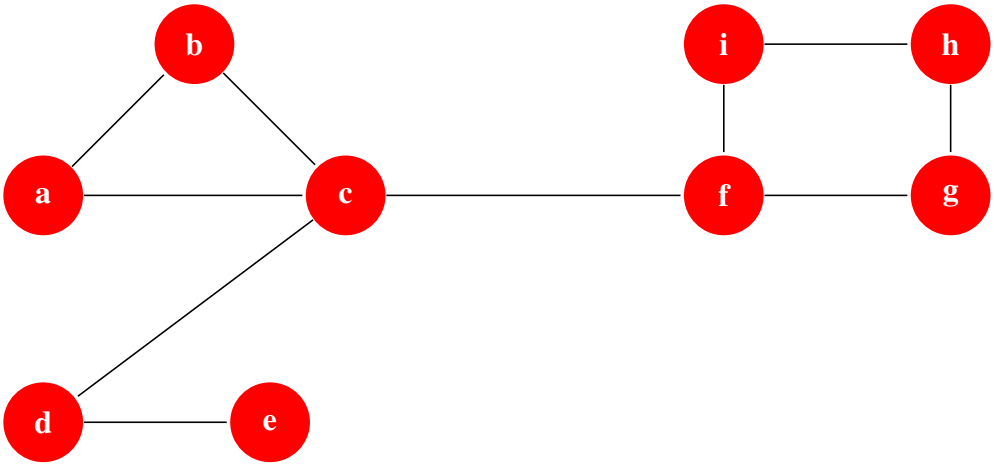


a) Determine the betweenness for each edge of G, considering each vertex as a starting point.

b) Determine the betweenness for G when only considering $s = \{a, d, e\}$ as starting points (sampling). Explain your observations compared to subtask a)! Would other instances of s lead to notably different results?

Exercise 4: Modularity (homework)

Given is the following graph G:



Given are two ways to partition this graph G:

- $s_1 : \{a, b, c\}$, $s_2 : \{d, e\}$, $s_3 : \{i, h, f, g\}$
- $s'_1 : \{a, b, c, d, e\}$, $s'_2 : \{i, h, f, g\}$

Compute the **modularity** for both partitionings and decide which one is better.