

Knowledge Discovery in Databases II
 WS 2012/2013

Übungsblatt 10: Graph Daten

Aufgabe 10-1 *Graphlet Kernels*

Eine elegante Möglichkeit die Topologie eines Graphen in einem Vektorraum darzustellen sind Graphlets. Ein Graphlet ist dabei ein Subgraph mit genau k Knoten. Jedes k -Graphlet unterscheidet sich dabei durch seine Topologie d.h. die Anordnung der Kanten. Die Vektordarstellung des Graphen $G(V, E)$ wird nun über die Auftrittshäufigkeiten der einzelnen k Graphlets in $G(V, E)$ gebildet.

- (a) Wie viele Graphlets der Größe 4 sind in einem Graphen mit n Knoten enthalten? Wie groß ist der Zeitaufwand, um sie zu bestimmen?
- (b) Wie teuer ist der Vergleich zweier Graphen mittels Graphlets der Größe 4?
- (c) Wie kann man den Vergleich der Graphletvektoren beschleunigen/verbessern?
- (d) Welche Probleme treten für Größere Werte von k auf? Welche Konsequenzen treten bei einem knoten-gelabelten Graphen $G(V, E, L)$ auf?

Aufgabe 10-2 *Shortest Path Kernel und Wiener Index*

Wir vergleichen zwei Graphen mittels des Shortest Path Kernel.

Außerdem sei der Wiener Index $W(G)$ eines Graphen G gegeben als:

$$W(G) = \sum_{v_i \in G} \sum_{v_j \in G} d(v_i, v_j),$$

wobei $d(v_i, v_j)$ die Länge des kürzesten Pfades zwischen den Knoten v_i und v_j aus G ist.

Welcher Zusammenhang besteht zwischen dem Wiener Index und dem Shortest Path Kernel?

Aufgabe 10-3 *Canonical Labeling*

Berechnen Sie das Canonical Labeling für folgende Graphen:

