

## Knowledge Discovery in Databases II

Wise 11/12

### Übungsblatt 09: Graph Mining

Besprechung am 20.01.12

#### Aufgabe 09-1 *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.

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

#### Aufgabe 09-2 *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:

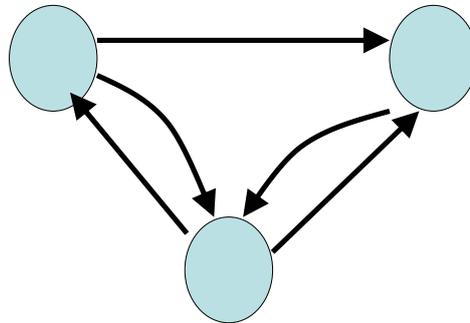
$$W(G) = \sum_{v_i \in G} \sum_{v_j \in G} d(v_i, v_j),$$

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 09-3 (Pagerank)**

Gegeben sei folgender (Web-)Graph.



- (a) Bestimmen Sie die Adjazenzmatrix.
- (b) Bestimmen Sie die Matrix  $L$  der Übergangswahrscheinlichkeiten für das Modell des Zufallssurfers bei PageRank.
- (c) Bestimmen Sie die Aufenthaltswahrscheinlichkeiten des Zufallssurfer zum Zeitpunkt  $t = 5$ . Als Startverteilung soll eine Gleichverteilung auf allen Knoten angenommen werden.
- (d) Bestimmen Sie jetzt die stationäre Aufenthaltswahrscheinlichkeiten für  $t = \infty$ .